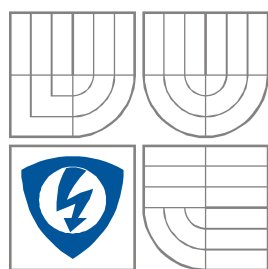


VYSOKÉ UČENÍ TECHNICKÉ V
BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A
KOMUNIKAČNÍCH
TECHNOLOGIÍ

ÚSTAV RADIOELEKTRONIKY

FACULTY OF ELECTRICAL ENGINEERING AND
COMMUNICATION

DEPARTMENT OF RADIO ELECTRONICS

ZPRACOVÁNÍ ŘEČOVÝCH SIGNÁLŮ V ČASOVÉ OBLASTI

SPEECH SIGNAL PROCESSING IN TIME DOMAIN

SEMESTRÁLNÍ PROJEKT

SEMESTRAL PROJECT

AUTOR PRÁCE

Ján Marko

AUTHOR

VEDOUCÍ PRÁCE

prof. Ing. Milan Sigmund, CSc.

SUPERVISOR

BRNO, 2012



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav radioelektroniky

Bakalářská práce

bakalářský studijní obor
Elektronika a sdělovací technika

Student: Ján Marko

ID: 138958

Ročník: 3

Akademický rok: 2013/2014

NÁZEV TÉMATU:

Zpracování řečových signálů v časové oblasti

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s metodami zpracování řečových signálů v časové oblasti a stručně je popište. Zaměřte se zejména na určování znělosti řečových úseků a základní periody. Zpracujte rešerši dostupných publikací pojednávajících o této problematice. Vybrané algoritmy ověřte na reálném řečovém signálu. Vytvořte a ověřte autonomní program na určování parametrů řečových signálů v časové oblasti (periodicita, jitter, shimmer). Pořídte vlastní databázi vhodných nahrávek, vytvořeným programem zpracujte data z této databáze a získané výsledky statisticky vyhodnoťte.

DOPORUČENÁ LITERATURA:

[1] PSUTKA, J., MÜLLER, Z., MATOUŠEK, J., RADOVÁ, V. Mluvíme s počítačem česky. Praha: Academia, 2006.

[2] SIGMUND, M. Rozpoznávání řečových signálů. Skriptum. Brno: FEKT, VUT v Brně, 2007.

Termín zadání: 10.2.2014

Termín odevzdání: 30.5.2014

Vedoucí práce: prof. Ing. Milan Sigmund, CSc.

Konzultanti bakalářské práce:

doc. Ing. Tomáš Kratochvíl, Ph.D.

Předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Táto bakalárska práca sa zaoberá metódami spracovávania rečového signálu. Práca obsahuje rešerš dostupných publikácií v problematike určovania znelosti úsekov reči a základnej frekvencie. Podrobne sa venuje metódam na rozpoznávanie vlastností reči v časovej oblasti. Porovnávaním teoretických a praktických výsledkov na reálnom rečovom signále vyhodnocuje použitie metód, poukazuje na ich výhody a nevýhody.

KLÚČOVÉ SLOVA

Fundamentálna frekvencia, periodičita, metóda krátkodobej energie, funkcia počtu prechodov nulou, krátkodobá autokorelačná funkcia v časovej oblasti, jitter, shimmer.

ABSTRACT

The bachelor's project deals with the processing of speech signals. The work includes a search of available publications on the issue of determining the voiced segments of speech and fundamental frequency. Attention is devoted to methods for speech recognition features in the time domain. Comparing theoretical and practical results on the real speech signals evaluates the use of methods, highlighting their advantages and disadvantages.

KEYWORDS

Fundamental frequency, periodicity, method of short-term energy, function of the zero crossing number, short-term autocorrelation function in the time domain, jitter, shimmer.

MARKO, J., *Zpracování řečových signálů v časové oblasti*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Ústav radioelektroniky, 2013. 53 s. Bakalářská práce. Vedoucí práce: prof. ing. Milan Sigmund, CSc.

PROHLÁŠENÍ

Prohlašuji, že svou semestrální práci na téma Zpracování řečových signálů v časové oblasti jsem vypracoval samostatně pod vedením vedoucího semestrální práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené semestrální práce dále prohlašuji, že v souvislosti s vytvořením této semestrální práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne

.....

(podpis autora)

PODĚKOVÁNÍ

Děkuji vedoucímu bakalářské práce prof. ing. Milanovi Sigmundovi, CSc. za účinnou metodickou, pedagogickou a odbornou pomoc a další cenné rady při zpracování mé bakalářské práce.

V Brně dne

.....

(podpis autora)

OBSAH

Zoznam obrázkov	viii
Zoznam tabuliek	x
1 Úvod	1
2 Rešerš dostupných publikácií zaoberajúcich sa spracovávaním rečových signálov	2
2.1 Metody detekce základního tónu řeči.....	3
2.2 PSUTKA, J., MÜLLER, Z., MATOUŠEK, J., RADOVÁ, V. Mluvíme s počítačem česky. Praha: Academia, 2006	4
2.3 Pitch Extraction and Fundamental Frequency:	5
2.4 Accurate short – term analysis of the fundamental frequency and the harmonics – to – noise ratio of a sampled sound	6
2.5 Jitter and Shimmer Measurements for Speaker Recognition	7
3 Tvorba reči a jej zachytávanie	8
3.1 Vznik reči	8
3.2 Proces vytvárania znělých a neznělých částí řeči	11
3.3 Delenie hlások.....	12
4 Metódy spracovania rečového signálu	14
4.1 Rozdelenie metód rozpoznávania řeči.....	14
4.2 Metódy spracovávania řeči v časovej oblasti	15
4.2.1 Analýza krátkodobé energie	17
4.2.2 Krátkodobá funkcia prechodu signálu nulou	18
4.2.3 Krátkodobá autokorelačná funkcia.....	21
4.2.4 Metóda detekcie vrcholov.....	21
5 Realizácia aplikácie	23
5.1 Dôvody použitia metód	23
5.1.1 Krátkodobá autokorelačná funkcia.....	23
5.1.2 Metóda detekovania vrcholov	25
5.1.3 Spoločná metóda	26
5.1.4 Kontrolná metóda.....	28
5.1.5 Výsledná metóda.....	28

5.2	Parametre aplikácie	28
5.2.1	Objem vstupných dát	28
5.2.2	Segment, mikrosegment, prah, vzdialenosť vrcholov	28
5.2.3	Rozsah detekovateľných základných frekvencií	29
5.2.4	Jitter, shimmer	29
6	Štatistické údaje	30
6.1	Jeden rečový záznam spracovaný všetkými metódami	30
6.2	Databáza záznamov spracovaných všetkými metódami	34
6.2.1	Rovnaký rečník	34
6.2.2	Rovnaký text v zázname naprieč databázou	37
6.2.3	Rôzne rečové módy: čítanie, šepot, rozprávanie, spev	37
7	Záver	40
	Literatúra	42
	Zoznam symbolov, veličín a skratiek	43
	Príloha A	44

ZOZNAM OBRÁZKOV

Obr. 3.1 Základná štruktúra dýchacích ciest (prevzaté z [4]).....	9
Obr. 3.2 Bloková schéma dýchacích ciest	9
Obr. 3.3 a) Hlasivky pri voľnom priechode vzduchu, b) hlasivky pri tvorbe reči (prevzaté z [6]).....	10
Obr. 3.4 časti hlasového a artikulačného ústrojenstva (prevzaté z [5]).....	11
Obr. 3.5 Prvá a druhá formantová frekvencia písmena A v Matlabe	11
Obr. 3.6 Vokálny trojuholník	12
Obr. 4.1 Pravouhlé okienko, jeho priebeh (hore), spektrum vypočítané pomocou FFT(v strede), posunuté spektrum pre zvýraznenie hlavného laloku (dole).....	16
Obr. 4.2 Hammingovo okienko, jeho priebeh (hore), spektrum vypočítané pomocou FFT (v strede), posunuté spektrum pre zvýraznenie hlavného laloku (dole)	17
Obr. 4.3 Priebeh vysloveného slova „filmov“ (hore), signál analyzovaný pomocou krátkodobej energie (v strede) a krátkodobej intenzity (dole).....	18
Obr. 4.4 Priebeh neznej spoluhlásky „c“ s vypočítaným počtom prechodov nulou	19
Obr. 4.5 Priebeh neznej spoluhlásky „s“ s vypočítaným počtom prechodov nulou	20
Obr. 4.6 Priebeh znej samohlásky „i“ s vypočítaným počtom prechodov nulou.....	20
Obr. 4.7 Priebeh znej samohlásky „o“ s vypočítaným počtom prechodov nulou	21
Obr. 5.1 Priebeh signálu v predĺženom pravouhlom okienku (hore) a vyhladený priebeh signálu po spracovaní modifikovanou autokorelačnou funkciou (dolu) so zvýrazneným maximom	24
Obr. 5.2 Celý signál (hore) so základnou frekvenciou okolo 100 Hz, vypočítaná základná frekvencia len pomocou metódy detekovania vrcholov (dolu).....	25
Obr. 5.3 Použitie detekcie vrcholov na reálnom signále, zachytená nepresnosť pri detekcii vrcholov v rečovom signále s nízkou základnou frekvenciou.....	26
Obr. 5.4 Výstup autokorelačnej metódy aplikovanej na dva segmenty so zvýrazneným prahom a detekovnými vrcholmi	27
Obr. 6.1 Graf celého signálu (vľavo hore), hodnoty základnej periódy vzhľadom na poradie segmentu po spracovaní modifikovanou autokorelačnou metódou (vpravo hore), metódou detekcie vrcholov (vľavo dole) a spoločnou metódou (vpravo dole)	32
Obr. 6.2 Graf celého signálu (vľavo hore), hodnoty základnej periódy vzhľadom na poradie segmentu po spracovaní kontrolnou metódou (vpravo hore), výstupnou založenou na spoločnej (vľavo dole) a výstupnou založenou na kontrolnej (vpravo dole)	32

- Obr. 6.3 Histogramy jitteru po spracovaní modifikovanou autokorelačnou metódou (vľavo hore), detekciou vrcholov (vpravo hore), spoločnou metódou (vľavo dole) a kontrolnou metódou (vpravo dole)33
- Obr. 6.4 Histogramy jitteru po spracovaní výslednou metódou založenou na spoločnej metóde (vľavo) a výslednou metódou založenou na kontrolnej metóde (vpravo)33

ZOZNAM TABULIEK

Tab. 1 Rozdelenie slovenských spoluhlások	12
Tab. 2 Porovnanie požiadaviek na výpočtovú techniku autokorelačnou a modifikovanou autokorelačnou funkciou	24
Tab. 3 Vypočítané výstupné parametre aplikácie pre jeden rečový záznam	30
Tab. 4 Vypočítané výstupné parametre aplikácie pre rovnakého rečníka (žena)	35
Tab. 5 Vypočítané výstupné parametre aplikácie pre rovnakého rečníka (muž)	36
Tab. 6 Vybrané výstupné parametre po spracovaní záznamov s rovnakým textom	37
Tab. 7 Vybrané výstupné parametre po spracovaní 5 záznamov s rečovým módom čítanie	38
Tab. 8 Vybrané výstupné parametre po spracovaní 5 záznamov s rečovým módom šepot	38
Tab. 9 Vybrané výstupné parametre po spracovaní 5 záznamov s rečovým módom rozprávanie	39
Tab. 10 Vybrané výstupné parametre po spracovaní 5 záznamov s rečovým módom spev	39

1 ÚVOD

Ústna forma informácií nemá dlhú životnosť, preto je snaha o premenu, v tomto prípade sa jedná konkrétne o transkripciu do písomnej formy. Táto práca hľadá optimálny spôsob spracovávania rečových signálov na základe štúdií dostupných materiálov a podkladov v tlačenej, prípadne elektronickej forme. V týchto zdrojoch sa nachádzajú informácie o metódach, ich aplikácii na signál, výhody a nevýhody ich použitia, požiadavky na výpočtovú techniku a mnoho iných informácií, ktoré poskytujú dostatočne objektívny pohľad na túto problematiku. Takto nadobudnuté poznatky je dobré využiť pri tvorbe skutočného funkčného algoritmu a tak porovnávať praktické výsledky a teoreticky nadobudnuté znalosti v tejto téme.

Hlavné teórie, ktoré riešia problematiku spracovania rečových signálov sú založené na poznatkoch o ľudskom tele, o jeho činnosti, o orgánoch a ústrojenstvách, ktoré sa podieľajú priamo alebo nepriamo na tvorbe rečového signálu, aký vplyv majú rozmery orgánov, čo spôsobuje zmena emocionálneho stavu hovoriaceho a mnohé iné fyziologické procesy. Taktiež berú ohľad na vplyv prostredia, ktoré môže spôsobovať šum, odrazy zvuku, vzdialenosť zdroja zvuku a prijímača, vzorkovanie signálu a s tým spojená digitalizácia. Zvuk je totiž spojitý v čase, má náhodný – stochastický charakter, šíri sa na základe princípu, podľa ktorého sa každý bod do ktorého sa zvuková vlna dostala, stáva novým elementárnym zdrojom zvuku. Zvuk je mechanické vlnenie, preto potrebuje pre svoj vznik častice, ktoré môžu kmitať.

Pre pochopenie tejto tématiky si treba uvedomiť, že reč rovnako ako zvuk vzniká šírením pulzov s väčšou a menšou hustotou vzduchu. Ak je toto striedanie periodické, možno povedať, že sa jedná o znelý úsek reči a teda má špecifické vlastnosti, ktoré sa dajú ďalej skúmať a využívať na rozpoznávanie.

2 REŠERŠ DOSTUPNÝCH PUBLIKÁCIÍ ZAOBERAJÚCICH SA SPRACOVÁVANÍM REČOVÝCH SIGNÁLOV

Štúdiom nasledujúcich publikácií možno nadobudnúť základný prehľad o tematike spracovania rečových signálov. Sú v nich riešené otázky ako určiť periodicitu signálu, ako nájsť a určiť fundamentálnu frekvenciu a akú metódu je na to vhodné použiť. Tiež sú rozoberané metódy, ktoré používajú odlišné procesy na činnosť so zvukovým signálom (napríklad ho vytvárajú), preto možno získať aj širší obraz o tom, aké je ďalšie použitie získaných informácií.

Ďalej riešia problémy pri určovaní základných parametrov spôsobené šumom prostredia, rôznorodosťou hovoriacich, vplyvom formantových frekvencií a iných. Preto vzniklo veľké množstvo metód, ktoré sú od prvotných odlišené napríklad použitím iného váhového okienka alebo je signál filtrovaný.

Navyše sú zdrojom potrebných vzťahov pre výpočet základných parametrov rečového signálu. Pre túto prácu je prínosom hlavne výpočet základnej periódy rôznymi metódami a absolútnych a relatívnych hodnôt jitteru a shimmeru.

2.1 Metody detekce základního tónu řeči

Ing. Hicham Atassi

Ústav telekomunikací, FEKT VUT, Purkyňova 118, Brno

Email: hicham.atassi@phd.feec.vutbr.cz

Tento příspěvek je zaměřený, ako z názvu vyplýva, na metody, které sú schopné detekovat přítomnost a frekvenci základního tónu řeči. Delia ich na tri základné oblasti, v ktorých sa fundamentálna frekvencia vyhľadáva. Konkrétne na časovú oblasť, frekvenčnú oblasť a kepstrálnu oblasť.

Metódy a ich výsledky, ktoré sú v tomto diele spracované sú autokorelačná, metóda centrálného klipovania, metóda inverznej filtrácie, spektrálna metóda, kepstrálna metóda. Pre každú metódu sa tu nájdú výsledky simulácií, ktoré sú porovnané a zachytené do spoločných grafov so zameranými hodnotami.

Ďalej sa tu nachádza vyhodnotenie, ktoré zoraďuje metódy v závislosti na presnosti, odolnosti na šume a náročnosti na výpočtovú techniku. Podľa týchto výskumov sa ako najpresnejšia metóda javila metóda centrálného klipovania, ktorá ku svojej korektnej činnosti využíva prah, ktorý sa vypočítava osobitne pre každý segment.

Tento příspěvek je možno nájsť na webovej adrese:
<http://www.elektrorevue.cz/cz/clanky/zpracovani-signalu/40/metody-detekce-zakladniho-tonu-reci/>

2.2 PSUTKA, J., MÜLLER, Z., MATOUŠEK, J., RADOVÁ, V. Mluvíme s počítačem česky. Praha: Academia, 2006

Toto knižné spracovanie poznatkov v oblasti rozpoznávania reči bližšie rozoberá procesy vytvárania reči, zohľadňuje vznik zvuku podľa miesta a spôsobu artikulácie, a tak zatrieduje rečové jednotky do skupín a fonetických inventárov. Nevýhodou je, že celá kniha je zameraná na delenie českých rečových jednotiek jazyka, preto bolo treba niektoré údaje kontrolovať a prípadne aj upraviť pre slovenský jazyk. Slovenský jazyk je však podobný, odlišuje sa len v počte foném a nie v akustickej tvorbe hlások, preto možno použiť rovnakú analýzu a metódy na spracovanie rečového signálu.

Vďaka svojmu rozsahu odкрýva základné vlastnosti rečových signálov, ako farba či výška hlasu, intonácia, prízvuk alebo rozdiel medzi spontánnou rečou a čítaním textu. Zaoberá sa problémami, s ktorými sa musia jednotlivé metódy pre korektnú činnosť vysporiadať ako je kvalita signálu, šum, ich odstup, vznik formantových frekvencií

Najdôležitejšou kapitolou pre túto prácu je analýza rečového signálu, kde sú veľmi presne rozpísané metódy schopné rozoznávať zvuk, od dôvodov, prečo deliť zachytený signál na mikrosegmenty, cez voľbu a použitie okienok, po analýzu získaných priebehov. Z nich možno určovať reč alebo ticho, znelé a neznelé časti reči, hodnotu základnej frekvencie hlasu. Všetky parametre sa dajú určovať na základe zvolenej metódy. Každá má svoje výhody aj nevýhody, je zameraná na inú vlastnosť, ktorú je schopná detekovať.

V tejto publikácii je aj veľké množstvo informácií, ktoré neboli priamym prínosom pre túto prácu, avšak osvetlili činnosti zariadení pracujúce s rečovým signálom – napríklad telefónne zariadenia, či ďalšie využívanie získaných dát.

2.3 Pitch Extraction and Fundamental Frequency:

History and Current Techniques

(Extrahovanie vrcholov a základnej frekvencie: Historické a súčasné techniky)

David Gerhard
Technical Report TR-CS 2003-06
November, 2003
c David Gerhard
Department of Computer Science
University of Regina
Regina, Saskatchewan, CANADA
S4S 0A2
ISSN 0828-3494
ISBN 0 7731 0455 0

Problematika tejto publikácie je zameraná na tri metódy, ktoré zachytávajú vrcholy intenzity a hodnoty základnej frekvencie. Porovnáva ich a na základe zlyhaní a usudzuje, ktorá metóda je najpresnejšia.

Taktiež upozorňuje na rozdiely detekcie v speve alebo v šepkaní, kde sú intenzity výrazne odlišné od spontánnej reči. Extrahovanie vrcholov je založené na princípe prahu, ktorý je dynamický, a teda je špecifický pre rôzne druhy zachyteného signálu, ktorému sa prispôbuje.

Pozornosť sa venuje aj rozdielom zvuku s rôznou intenzitou, zachytenému sluchom človeka alebo elektronickým prijímačom (mikrofónom), teda krivkám rovnakej hlasitosti, vplyvom maskovania frekvencií a zotavovania sluchových orgánov.

Zdôrazňuje náročnosť porovnávaní detektorov základnej frekvencie práve kvôli tomu, že každá aplikácia môže byť zameraná na inú oblasť reči alebo spevu. Tiež je dôležité poznamenať, že meranie základnej frekvencie je do dnes náročný proces a preto je problém určiť správnu hodnotu základnej frekvencie, s ktorou by sa mohli získať výsledky porovnávať. Používa sa preto metóda, pri ktorej sa zmeria spoločný súbor hodnôt F_0 a porovná sa s hodnotou, ktorá bola detekovaná najviac krát.

Metódy zmienené v tejto správe spracovávajú signál v časovej, frekvenčnej alebo statickej frekvenčnej oblasti. Táto práca je však zameraná na spracovávanie v časovej oblasti, preto sú použiteľné informácie o metódach prechodu signálu nulou (zero crossing rate), vzdialenosť vrcholov (peak rate) a funkcia hrán (slope event rate). Vzdialenosť lokálnych maxím udáva dĺžku periódy, prevrátená hodnota tejto periódy je základná frekvencia.

Túto publikáciu možno nájsť aj na webovej adrese: http://scholar.googleusercontent.com/scholar?q=cache:NeLo8MDPZB0J:scholar.google.com/+pitch+frequency+fundamental+period+metod&hl=sk&as_sdt=0&as_vis=1

2.4 Accurate short – term analysis of the fundamental frequency and the harmonics – to – noise ratio of a sampled sound

(Presná krátkodobá analýza základnej frekvencie a pomer harmonických funkcií ku šumu na vzorkovanom zvuku)

Paul Boersma

Institute of Phonetic Sciences,

University of Amsterdam,

Proceedings 17 (1993), 97-110.

Tento článok sa venuje autokorelačnej metóde, pomocou ktorej možno rozpoznať periodicitu – znelosť zvukového záznamu, druhom okienok, dôvodom ich použitia, ich výhody a nevýhody, časovému oneskoreniu, vzorkovaniu signálu, využívaniu rýchlej Fourierovej transformácie (FFT), základné podmienky jej správneho fungovania a čo možno vyčítať z takto vytvoreného spektra.

Tieto metódy sú testované na signáloch, ktoré sú čisto periodické, čiastočne periodické, neperiodické, s výrazným odstupom šumu a signálu (v článku označované skratkou HNR – z anglického výrazu harmonic to – noise - ratio) aj porovnateľnej intenzite šumu a signálu. Nevenujú sa iba rečovým zvukom, tieto princípy sú totiž platné pre širší rozsah frekvencií od infrazvukov (pod 20 Hz) po detský krik (1800 Hz) a zahrňujú tak aj rôzne vibrácie, ktoré nepočujeme a frekvencie, ktoré nie sú pre určenie základnej frekvencie reči nutnosťou.

Takisto sa tu nachádzajú dva postupy ako vytvoriť algoritmus na spracovávanie rečového signálu pomocou autokorelačnej funkcie spojenej s okienkom a algoritmus s použitím FFT analýzy.

Záver z toho článku sú, že výsledky získané pomocou autokorelačnej metódy sú menej presné ako pri spracovávaní oneskoreného signálu pomocou FFT, aj keď naprogramovanie tohto algoritmu je oveľa náročnejšie a je treba uskutočniť viac krokov ku korektnému fungovaniu tejto analýzy.

Tento článok je dostupný na tejto webovej adrese :
http://xeds.eu/other/P_Boersma_Accurate_short-term_analysis_of_the_fundametnal_freq.pdf

2.5 Jitter and Shimmer Measurements for Speaker Recognition

(Merania jitteru a shimmeru na rozoznávanie rečníka)

Mireia Farrús, Javier Hernando, Pascual Ejarque

TALP Research Center, Department of Signal Theory and Communications

Universitat Politècnica de Catalunya, Barcelona, Spain

{mfarrus,javier,pascual}@gps.tsc.upc.edu

Táto správa pojednáva o niekoľkých možnostiach merania a spôsoboch výpočtu jitteru a shimmeru z fundamentálnej frekvencie. Navyše tieto pojmy – jitter a shimmer vysvetľuje. Jitter a shimmer patria medzi akustické informácie, ktoré je možné získať z rečových signálov. Sú vyčíslené ako variácia základnej frekvencie, respektíve tvaru vlny amplitúdy. Tieto vlastnosti sú veľmi závislé na emocionálnom stave rečníka a hlavne na štýle rozprávania.

Dôraz sa v tejto správe kladie na rozoznávanie rečníka práve pomocou vypočítavania jitteru a shimmeru naprieč celou databázou záznamov. Na výpočet základnej frekvencie sa používa autokorelačná metóda so vstupom, ktorý je upravený okienkom s dĺžkou 40/3 ms a porovnávanou časťou - mikrosegmentom dĺžky 10/3 ms. Postupným zlučováním a porovnávaním výsledkov pre rôzne kombinácie výpočtov sa zvyšuje úspešnosť rozoznávania rečníka. Najväčšia úspešnosť sa logicky považuje za najvhodnejšiu metódu, ktorej výstupné informácie sú použiteľné pre ďalšie spracovávanie.

Táto správa je dostupná na webovej adrese:
http://nlp.lsi.upc.edu/papers/far_jit_07.pdf

3 TVORBA REČI A JEJ ZACHYTÁVANIE

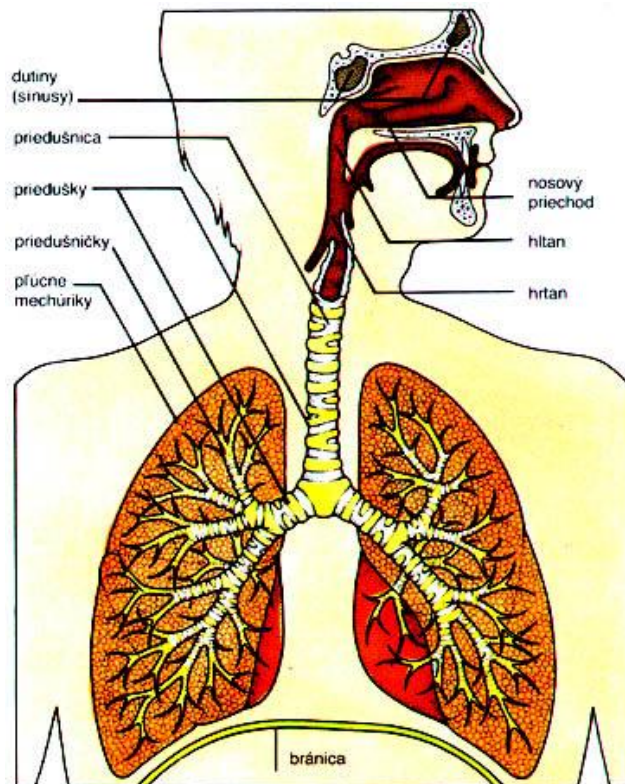
O vzniku reči sa dá polemizovať. Najpravdepodobnejšie je, že vznikla z nutnosti komunikácie medzi členmi rodiny, skupiny, tlupy, či kmeňa pri love. Začiatkové vymieňanie informácií prebiehalo prostredníctvom posunkovej reči, spevu a imitácii zvukov zvierat. Takto vznikol súbor znakov, symbolov a prvých slov, vďaka ktorým boli členovia schopní dorozumieť sa. Podľa jednej teórie sa táto slovná zásoba začala šíriť medzi ďalšie skupiny populácie a vznikla reč v ústnej forme. Predpokladá sa, že plne artikulovaný jazyk vznikol pred 70 000 rokmi. Niekoľko desiatok tisíc rokov sa ústna forma širila, zdokonaľovala, vznikali rozdiely, ktoré môžeme vidieť dodnes. Vznik písomnej formy reči sa podľa nájdených predmetov datuje na obdobie 3500 rokov pred Kristom [3].

Z toho možno usúdiť, že trvácnosť písma je väčšia, pri ústnom podaní sa správy šíria vzduchom, preto je ich trvácnosť malá. Intenzita klesá so štvorcom vzdialenosti od zdroja zvuku [8].

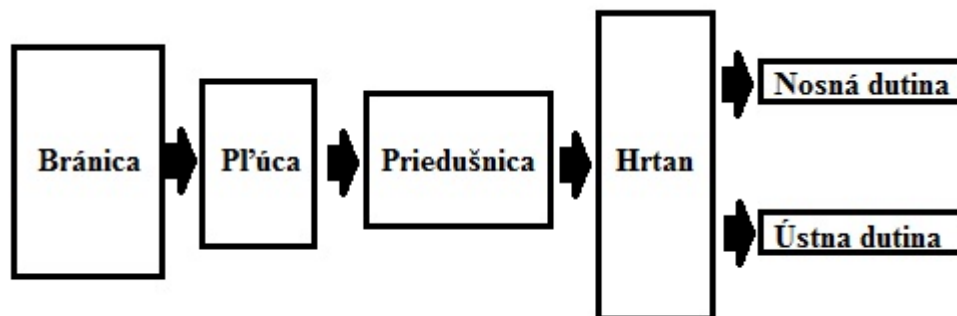
V písomnej forme boli nosičmi kamenné tabuľky, papyrus, koža, neskôr papier. V dnešnej dobe sa ukladajú data na elektronické zariadenia, kde sa teoreticky uchovávajú na nekonečne dlhú dobu a sú zdrojom informácií. To sa dnes využíva rôznym spôsobom, napríklad pri vytváraní referátov, blogov, článkov, seminárnych, bakalárskych alebo diplomových prác. Na dôležitosť písomnej formy ešte viac poukazuje vývoj výpočtovej techniky, ktorá je bežnou súčasťou domácností. Preto dnes text prevyšuje v niektorých oblastiach spoločenského života verbálnu komunikáciu, či už vo firmách zameraných na rozvoj informačných technológií alebo na sociálnych sieťach. Preto sú snahy o rozpoznávanie reči a jej transkripciu do písma, čo sa dá určite využiť aj pre zlepšenie životných podmienok nevidiacich ľudí. To by však nebolo možné bez základných poznatkov vzniku reči.

3.1 Vznik reči

Základným predpokladom tvorby reči je dýchanie. Je to biologický proces, ktorého primárnou funkciou je odvádzanie a vylučovanie oxidu uhličitého a vodnej pary (produktov látkovej premeny) z organizmu do prostredia a privádzanie kyslíku do pľúc. To všetko zabezpečuje dýchací trakt počas dvoch fáz –výdychu a nádychu. Tie sa cyklicky opakujú priemerne 20 krát za minútu v pokojovom stave organizmu. Dýchací trakt sa skladá z viacerých častí ako vidieť na obr. 3.1, ktoré sa súhrnne nazývajú dýchacie ústrojenstvo. Každá časť má dôležitú funkciu buď pre dýchanie alebo pre tvarovanie rečového signálu. Na obr. 3.2 je znázornená bloková schéma základných častí dýchacích ciest.



Obr. 3.1 Základná štruktúra dýchacích ciest (prevzaté z [4])



Obr. 3.2 Bloková schéma dýchacích ciest

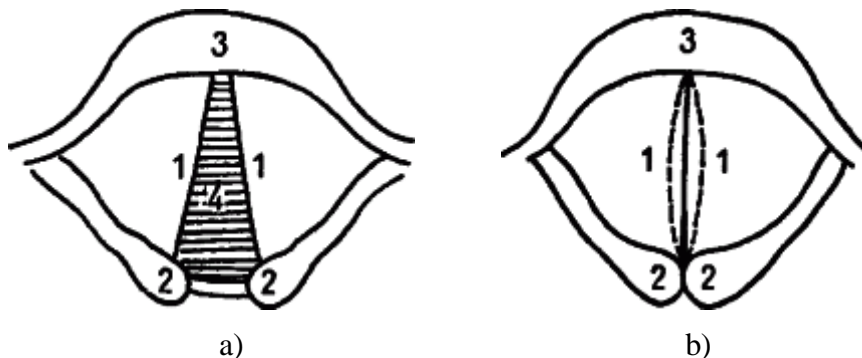
Bránica – hladký sval, ktorého činnosť nie je podmienená našim vedomím, pracuje reflexívne. Pri nádychu sa bránica sťahuje (kontrahuje), tým sa rozširuje hrudný kôš a do pľúc sa privádza vzduch z prostredia. Pri uvoľňovaní (relaxácii) a činnosti medzirebrových svalov nastáva výdych. Ten spôsobí, že vzduch získava kinetickú energiu.

Pľúca – tu nastáva látková výmena. Pľúca sú zložené z lalokov, ktoré sa nakoniec spoja do priedušnice a tá postupne prechádza do hrtanu. Objem pľúc je 3 až 5 litrov a je odlišný pre mužov a ženy. Počas výdychu pri tvorbe reči je množstvo odchádzajúceho vzduchu 0,2 litra za sekundu.

Hrtan – je to trubica, tvorená chrupavkami. Zmenou dĺžky môže meniť celkovú dĺžku dýchacích ciest a tak sa priamo podieľa na zmene zvuku. Avšak

výraznejšiu úlohu zohrávajú hlasivky (obr. 3.3), ktoré sú v ňom uložené. Hrtan je zakončený hlasivkovou príklopkou, ktorá chráni otvor do pľúc pri požívaní. Za ňou sa nachádza hrdlová dutina, ktorá sa ďalej člení na ústnu a nosnú.

Ústna dutina – začína sa špičkou mäkkého podnebia a končí perami. Od nosnej dutiny je oddelená v zadnej časti mäkkým, vpredu tvrdým podnebím, na okrajoch ktorého je vrchná polovica zubov. Druhá polovica zubov je spojená s čeľusťou a obklopuje jazyk. Tieto časti ovplyvňujú hlavne priebeh šumových zložiek hlasu.

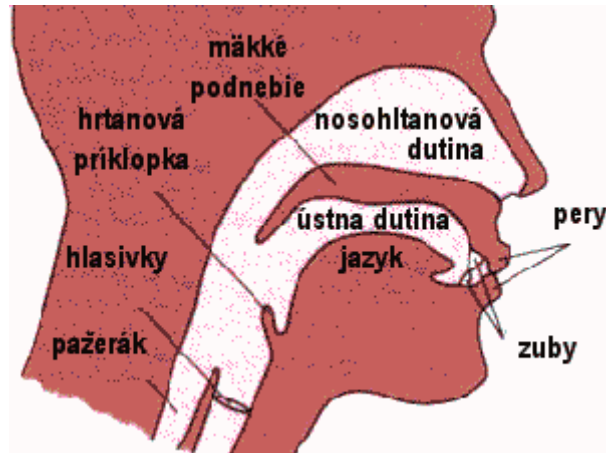


Obr. 3.3 a) Hlasivky pri voľnom priechode vzduchu, b) hlasivky pri tvorbe reči (prevzaté z [6])

Pre tvorbu reči z prúdenia vzduchu sa využívajú niektoré časti dýchacieho traktu spolu s prispôbenými orgánmi a súhrnne sa nazývajú ústrojenstvá. Delia sa na hlasové a artikulačné.

Hlasové ústrojenstvo – nachádza sa v hrtane. Hlavnou časťou sú hlasivky, ktoré sú spojené s hlasivkovou a štítnou chrupavkou. Sú tvorené sliznicou, hlasivkovými väzmi a svalmi. Záhyby sliznice vytvárajú riasy, hlasivkové väzy sú elastické, preto môžu meniť svoju veľkosť. Svaly sú priečne pruhované, a teda ich môžeme ovládať vôľou a ovplyvňovať tak šírku hlasivkovej štrbiny. Štrbina má štandardne plochu 20 mm^2 u mužov a 14 mm^2 u žien. Pri dýchaní má trojuholníkový tvar a je úplne otvorená. Pri vytváraní reči sa zužuje, vzduch naráža na hlasivkové riasy, ktoré vďaka svojej pružnosti začnú kmitať. Tým vznikajú miesta s väčšou a menšou hustotou vzduchu, ktoré sa takmer rovnomerne opakujú a majú teda periódu. Reč má preto pre každého človeka špecifickú základnú výšku závislú na veľkosti hlasiviek. Prevrátená hodnota tejto periódy sa nazýva frekvencia základného hlasivkového tónu a označuje sa F_0 . U mužov je táto frekvencia v rozsahu 80 – 160 Hz, výnimočne len 50 Hz, u žien 150 – 300 Hz, občas až 400 Hz, u detí 200 – 600 Hz. Hodnoty F_0 sú pri reči v rozmedzí jednej októvy. Spevom sa rozsah hlasu zväčšuje na dve až tri októvy. Výstupný rečový signál z hlasiviek je periodický, teda znělý, ak bola štrbina zúžená. [1]

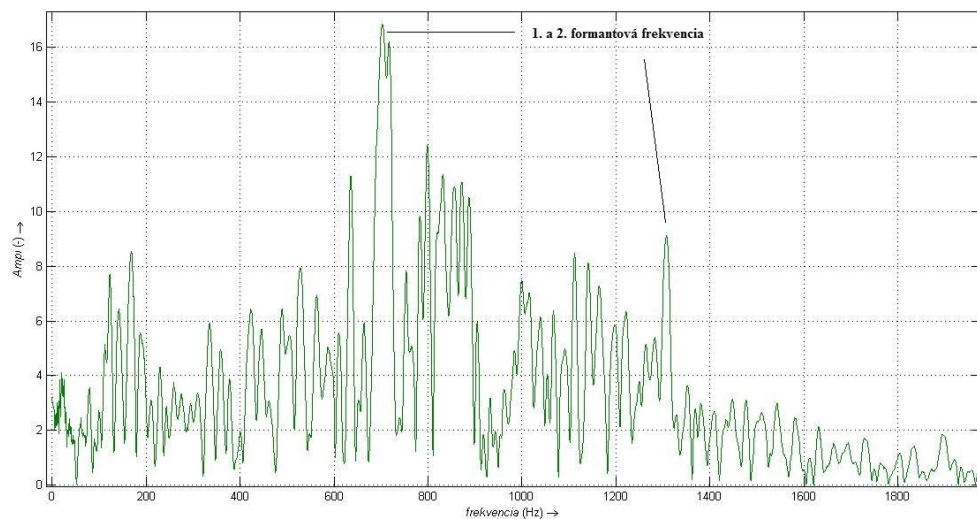
Artikulačné ústrojenstvo – je tvorené z viacerých častí, ktoré pasívne alebo aktívne tvarujú rečový signál. Medzi pasívne sa radia nepohyblivé časti. Sú to dutiny – hrdlová, ústna a nosná. Aktívne orgány menia rozmery a tvar dutín. Najväčšiu úlohu zohráva jazyk, ktorého časti sa dokážu meniť nezávisle a dostatočne rýchlo. Spolu s mäkkým podnebím a perami majú na svedomí veľkú rozmanitosť znělých aj neznelých hlások. Menšou časťou sa podieľajú hrtan, tvrdé podnebie, čeľuste a zuby.



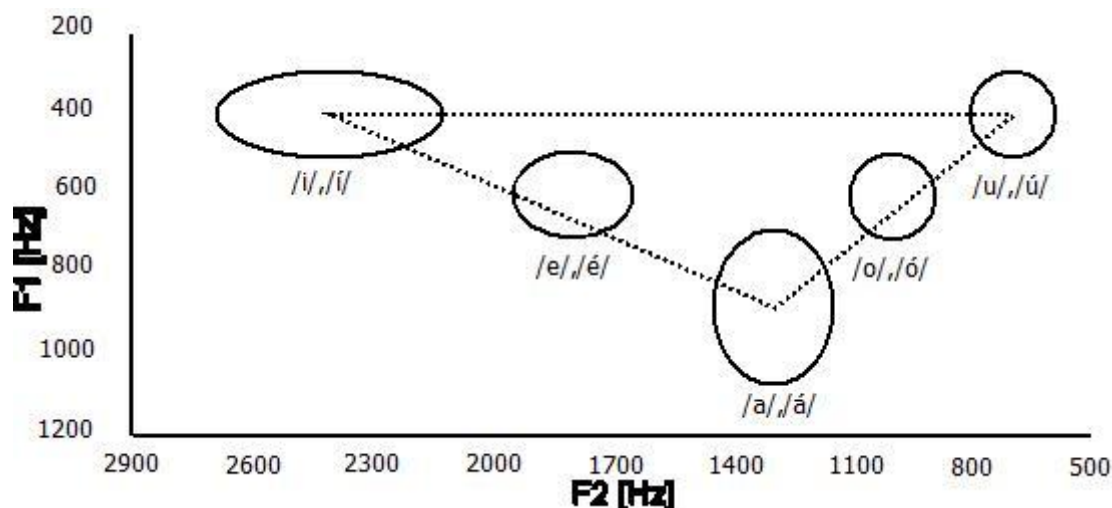
Obr. 3.4 časti hlasového a artikulačného ústrojenstva (prevzaté z [5])

3.2 Proces vytvárania znelých a neznelých častí reči

Prúd vzduchu prihádza do hrdlovej dutiny už s fundamentálnou periódou alebo voľne prešiel cez hlasové ústrojenstvo bez zmeny. Tu sa vplyvom rezonancie dutín mení energia na jednotlivých frekvenciách. Preto možno v znelom úseku reči pomocou rýchlej furierovej transformácie (FFT) ľahko sledovať vznik formantových frekvencií. Vďaka pomeru druhej a prvej formantovej frekvencie $F2/F1$, ktoré vidieť na obr. 3.5 možno určiť, akú samohlásku signál zachytáva. Tento spôsob je jednoduchý aj vďaka vokalicnému trojuholníku (obr. 3.6), avšak nie je veľmi presný.



Obr. 3.5 Prvá a druhá formantová frekvencia písmena A v Matlabe



Obr. 3.6 Vokálny trojuholník

Nevýhodou formantových frekvencií je narušanie metód pre získavanie základnej frekvencie, preto je často nutné signál najskôr upraviť pomocou filtra typu dolná priepusť, kde sa medzná frekvencia volí väčšia ako predpokladaná základná frekvencia, ale menšia ako formantové frekvencie (približne 400Hz). Tento spôsob je dobré použiť hlavne ak sa jedná o výpočtovo menej náročné spôsoby určovania znelosti, prípadne pitch frekvencie. Medzi tieto metódy patria krátkodobá analýza energetického alebo výkonového spektra, analýza spektra intenzity, funkcie prechodu signálu nulou a iné [1].

3.3 Delenie hlások

Reč je sled za sebou idúcich zvukov. Ak za sebou nasledujú podobné zvuky, nazývajú sa fonémy. Ak párové hlásky znejú podobne, ale význam vzniknutých slov je rôzny, hovoríme o dvoch fonémach. Napríklad *p* a *b* v slovách *puk* a *buk*. Hlásky sa delia na dve skupiny: samohlásky a spoluhlásky. Samohlásky sú tvorené periodickým signálom, preto sa radia medzi znelé. Spoluhlásky naopak pozostávajú zo šumu, prípadne sú sonorové. V závislosti na tom, z akej časti je signál šumový, delíme spoluhlásky na znelé a neznelé. Sonorové spoluhlásky sú nepárne, ale všetky sú znelé. Šumové sú párne – majú znelú aj neznelú fonému. Základné rozdelenie slovenských spoluhlások je znázornené v tab. 1.

Tab. 1 Rozdelenie slovenských spoluhlások

Párové	Neznelé	/p/	/t/	/tʰ/	/c/	/č/	/s/	/š/	/k/	/ch/	/f/
	Znelé	/b/	/d/	/dʰ/	/dz/	/dž/	/z/	/ž/	/g/	/h/	/v/
Nepárové	Znelé	/m/, /n/, /ň/, /l/, /j/, /r/									

Reč sa bežne zachytáva do digitálnej podoby pomocou mikrofónu. Zachytený signál v závislosti na veľkosti intenzity vytvára na prijímacom zariadení napätie.

Rečový signál je spojitý, avšak použiteľný vo výpočtovej technike je až po prevode na digitálny. Využíva sa na to vzorkovacia frekvencia. Označuje sa F_{vz} . Do digitálnej podoby sa prenášajú iba hodnoty v určitom čase. Aby zostala informácia použiteľná, je treba zachovať Niquistov vzorkovací teorém $F_{vz} \geq 2F_m$, ktorého základom je použitie minimálne dvojnásobnej vzorkovacej frekvencie, ako je frekvencia signálu F_m . Ak táto podmienka nie je splnená, dochádza ku skresleniu zložiek vyšších frekvencií [2].

Ľudským sluchom zachytávaný a počuteľný signál je do frekvencie 20 kHz. Preto sa často využíva vzorkovacia frekvencia 44 100 Hz, ktorá spĺňa vzorkovací teorém a navyše pokrýva celú počuteľnú frekvenčnú charakteristiku. Pre zachytávanie a prenášanie reči však nie je dôležitý tak široký rozsah, preto sa v telefónnych zariadeniach používajú mikrofóny zachytávajúce menší frekvenčný rozsah. Výhodou je, že nie je potrebná tak veľká vzorkovacia frekvencia a štandardom je 8 kHz, prípadne 16 kHz. Takto digitalizovaný signál má taktiež nižší počet zachytených prvkov, čo umožňuje vyššiu prenosovú rýchlosť. Vzorkovaním signálu vzniká pulzne modulovaný signál, ktorý je základnou podmienkou činnosti metód, zaoberajúcich sa spracovaním reči.

4 METÓDY SPRACOVANIA REČOVÉHO SIGNÁLU

Spracovaním rečového signálu sa zaoberá viacero inštitútov na Slovensku, v Českej Republike, ale aj po celom svete. Rôznorodosť metód ide ruka v ruku s technickým pokrokom, ktorý dáva nie len podnety, ale aj možnosti na spracovanie reči. Je snaha o úplnú transkripciu ústnej formy jazyka na písomnú, čo by podporilo vývoj techniky ovládanej ľudským hlasom. Avšak do dnešnej doby sa nepodarilo nájsť metódu, ktorá by bola schopná rozlišovať všetky fonémy a zároveň by nebola závislá na rečníkovi alebo naopak metóda, ktorá je nezávislá na osobe, ani na jej pohlaví, ale dokázala by rozlíšiť všetky fonémy. Vzhľadom na túto rozmanitosť jazykov a rečníkov, existuje veľké množstvo metód.

4.1 Rozdelenie metód rozpoznávania reči

Reč je veľmi komplikovaná kvôli svojej rozmanitosti, preto ju možno rozdeliť do niekoľkých tried podľa toho, čo spôsobuje jej odlišnosť. Zdroj rečového signálu má charakteristické vlastnosti v závislosti na hovoriacom, prípadne okolí, konkrétne:

- doba trvania
 - veľmi krátky segment (niekoľko minisekúnd trvajúca fonéma, prípadne hláska)
 - krátky rečový signál (niekoľko hlások, prípadne celé slovo)
 - dlhý rečový signál (slová, vety, súvetia v rozsahu maximálne niekoľkých minút)
 - veľmi dlhý záznam (teoreticky nekonečná dĺžka rečového signálu)
- amplitúda signálu
- výslovnosť
- rečový signál ovplyvnený aditívnym šumom (prostredím)

Metódy spracovávajúce reč sa zatriedujú podľa toho, aké podmienky splňujú:

Rečník:

- závislé – jeden rečník
- nezávislé – rôznorodosť rečníkov aj ich pohlaví

Rečový mód:

- slová oddelené pauzou
- plynulá reč

Spôsob prejavu:

- spontánna reč
- čítaný text

Veľkosť slovnej zásoby:

- malá (do 20 slov)
- stredná (20 až 20 000 slov)
- veľká (nad 20 000 slov)

Rozdiel signálu a šumu:

- nad 30 dB
- pod 10 dB

Kvalita a s tým súvisiace množstvo zachytených dát:

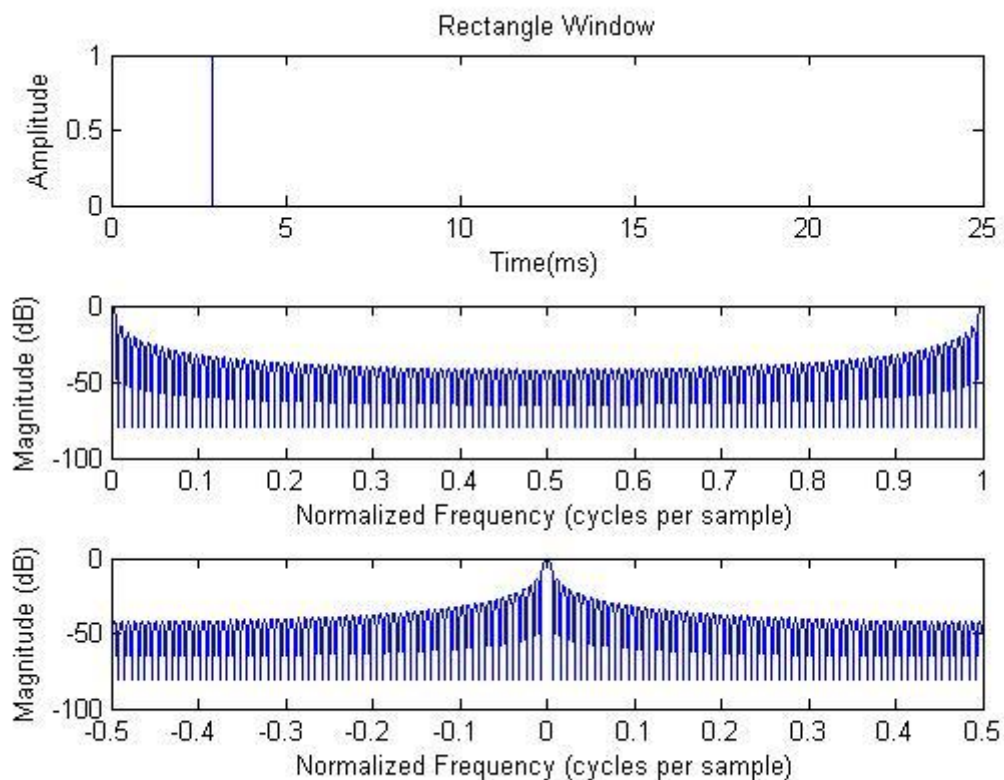
- telefónna kvalita – 8 000 vzoriek za sekundu
- kvalita audio nosičov – 44 100 vzoriek za sekundu
- profesionálne zvukové záznamy – nad 44 100 vzoriek za sekundu

4.2 Metódy spracovávaní reči v časovej oblasti

Prvý krok, ktorý treba urobiť pri spracovaní signálu je rozoznávanie zvuku od ticha. Túto informáciu môžeme získať jednoduchými metódami napríklad vďaka krátkodobej energii signálu. Pre využívanie tejto metódy, ale aj mnohých iných je treba signál upraviť pomocou tzv. okienka. Toto okienko zaručuje, že sa spracováva iba určitá časť signálu, zväčša o dĺžke jedného, prípadne niekoľkých segmentov, v závislosti na metóde a dĺžke zvoleného okienka. Najčastejšie sa používajú tieto dva druhy okienka: Pravouhlé okienko, ktoré je definované vzťahom

$$w(n) = \begin{cases} 1 & \text{pre } 0 \leq n \leq L-1 \\ 0 & \text{pre ostatné } n \end{cases}, \quad (4.1)$$

kde $n=Ni-1$, i je poradové číslo mikrosegmentu a L je počet vzoriek vybraných okienkom. Jeho spektrálna charakteristika je znázornená na obr. 4.1.

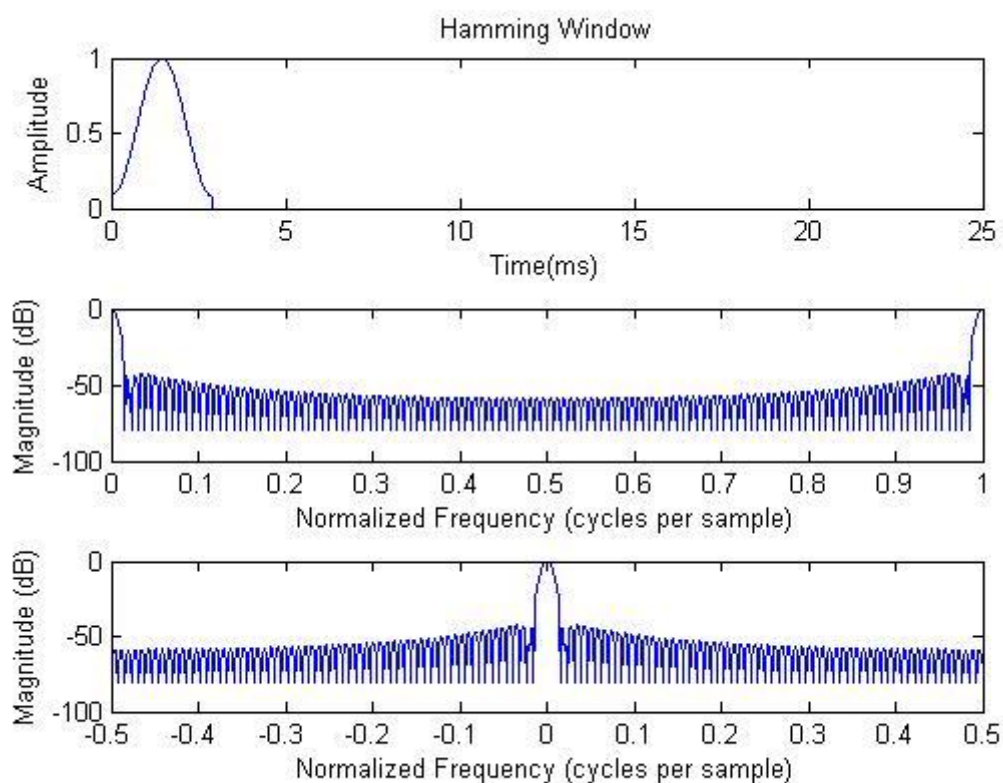


Obr. 4.1 Pravouhlé okienko, jeho priebeh (hore), spektrum vypočítané pomocou FFT (v strede), posunuté spektrum pre zvýraznenie hlavného laloku (dole)

Hammingovo okienko je popísané rovnicou

$$w(n) = \begin{cases} 0,54 - 0,46 \cos(2\pi n / (L-1)) & \text{pre } 0 \leq n \leq L-1 \\ 0 & \text{pre ostatné } n \end{cases} \quad (4.2)$$

Hammingovo okienko sa využíva hlavne pre väčší útlm a lepšiu priepustnosť nízkych frekvencií (obr. 4.2), čo je výhodné ak sa zo signálu ďalším spracovaním získava základná perióda, prípadne znelosť, lebo neznelé spoluhlásky majú na vyšších frekvenciách väčšiu energiu. Toto okienko taktiež čiastočne nahrádza filter typu dolný priepust. Medzi ďalšie typy okienok patria Hanningovo, Blackmanovo alebo trojuholníkové [7].



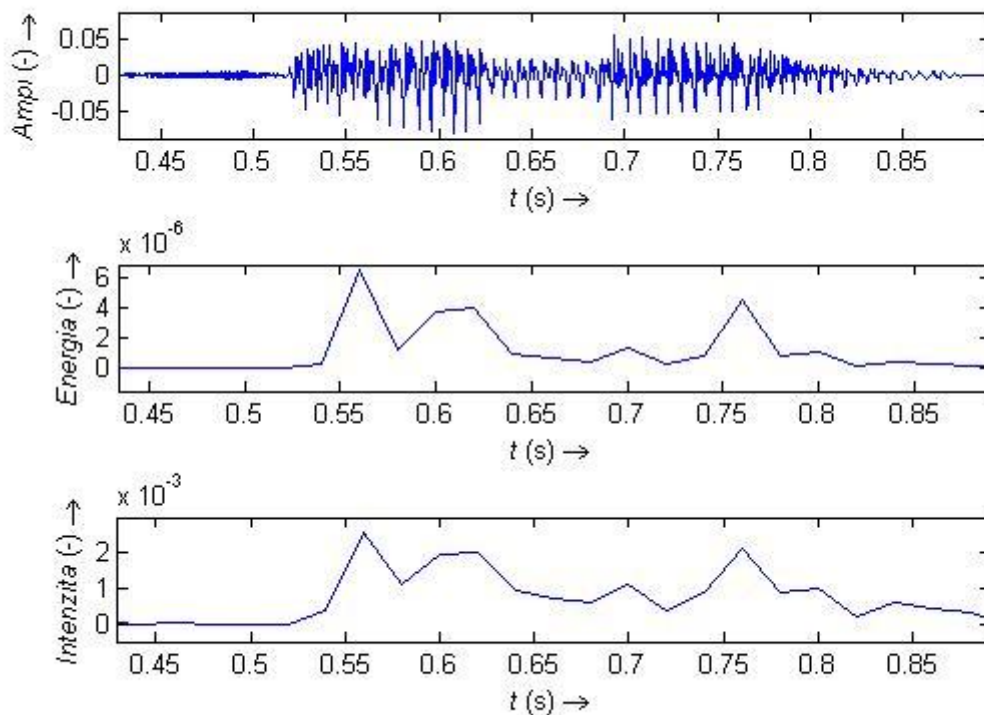
Obr. 4.2 Hammingovo okienko, jeho priebeh (hore), spektrum vypočítané pomocou FFT (v strede), posunuté spektrum pre zvýraznenie hlavného laloku (dole)

4.2.1 Analýza krátkodobej energie

Táto metóda sa používa na jednoduché spracovanie signálu, a to rozlišovanie ticha a reči, prípadne znělých a neznelých častí reči. Jej priebeh je zobrazený na obr. 4.3. Hodnoty boli vypočítané na základe

$$E_n = \sum_{k=-\infty}^{\infty} [s(k)w(n-k)]^2, \quad (4.3)$$

kde $s(k)$ je signál v čase k a $w(n)$ je okienko. Keďže je funkciou kvadratickou, je veľmi citlivá na prudké zmeny signálu.



Obr. 4.3 Priebeh vysloveného slova „filmov“ (hore), signál analyzovaný pomocou krátkodobej energie (v strede) a krátkodobej intenzity (dole)

Preto sa niekedy namiesto nej využíva radšej **metóda krátkodobej intenzity**, ktorá ako vidieť v jej definícii

$$M_n = \sum_{k=-\infty}^{\infty} |s(k)| w(n-k) \quad (4.4)$$

tento nedostatok nemá [1].

4.2.2 Krátkodobá funkcia prechodu signálu nulou

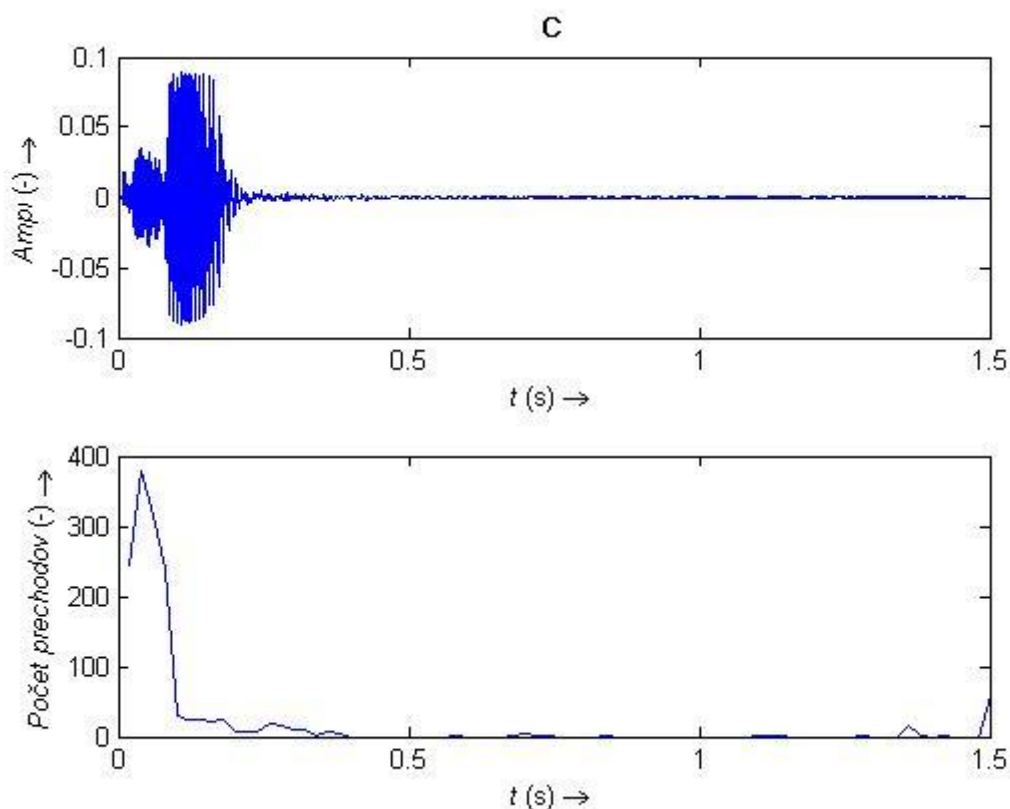
Ďalšia metóda na spracovanie reči je krátkodobá funkcia prechodu signálu nulou. Je definovaná ako

$$Z_n = \sum_{k=-\infty}^{\infty} |\operatorname{sgn}[s(k)] - \operatorname{sgn}[s(k-1)]| w(n-k), \quad (4.5)$$

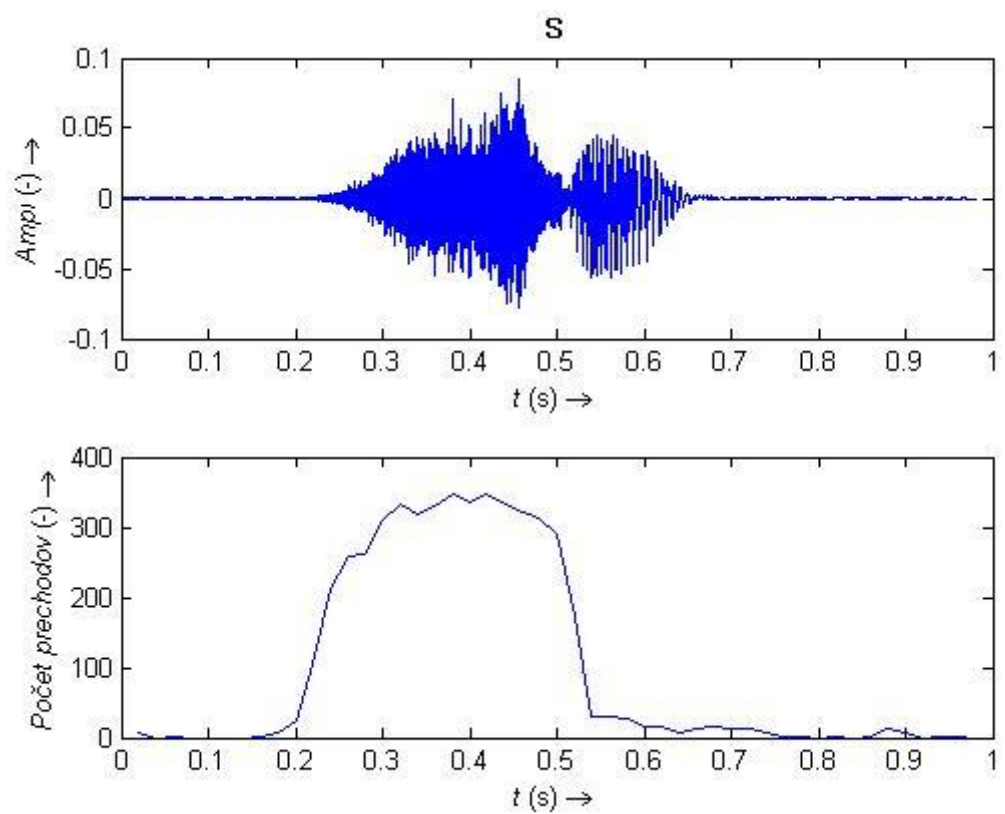
kde

$$\operatorname{sgn}[s(k)] = \begin{cases} 1 & \text{pre } s(k) \geq 0 \\ -1 & \text{pre } s(k) < 0 \end{cases} \quad (4.6)$$

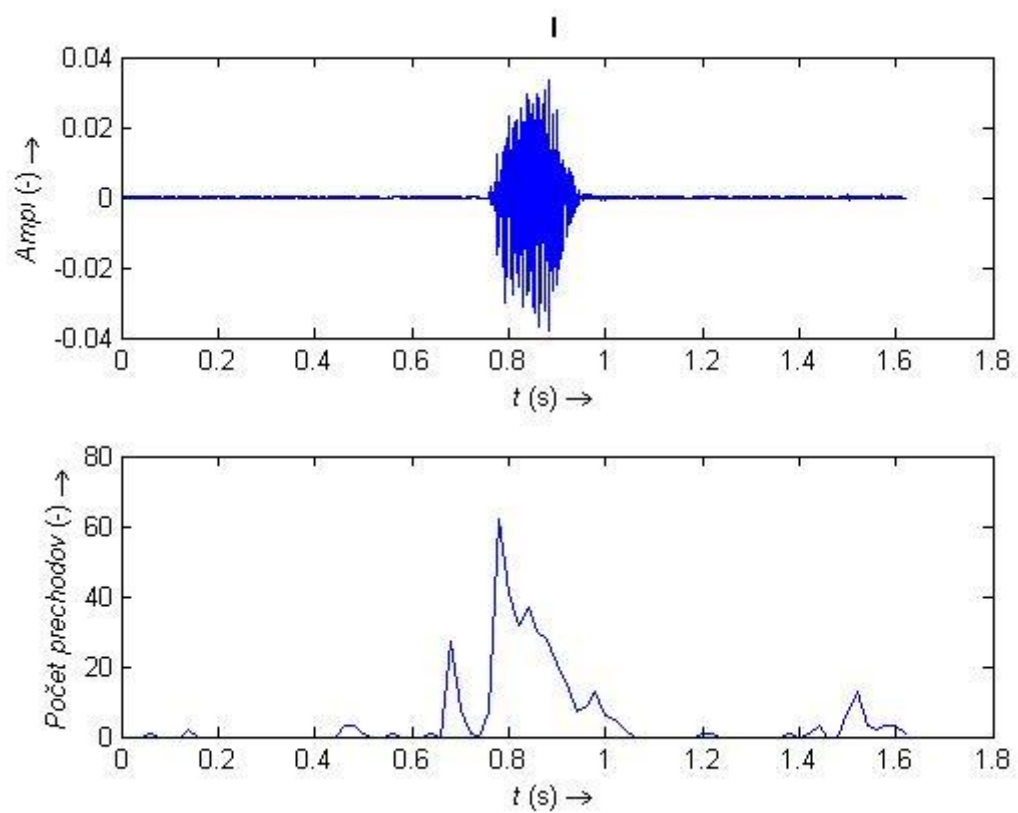
Uplatňuje sa a má široké využitie pri určovaní začiatkov a koncov slov, taktiež sa na základe jej priebehu dá určovať znelosť samohlások a spoluhlások, ktoré sa od neznelých líšia ako vidieť na obr. 4.4 výrazne nižším počtom prechodov nulou. Zachytený signál so získaným počtom prechodov je pre neznelé spoluhlásky „c“ a „s“ znázornený na obr. 4.5 a 4.6 a pre znelé samohlásky „i“ a „o“ na obr. 4.7 a 4.8. Na výpočet bol použitý vzorec (4.5) kde sa hodnota navyšuje po každom prechode nulou. Aby mohla byť analýza vykonávaná v reálnom čase, využíva sa aj u tejto metódy rozdelenie signálu na segmenty a ich upravenie pomocou zvoleného okienka [7].



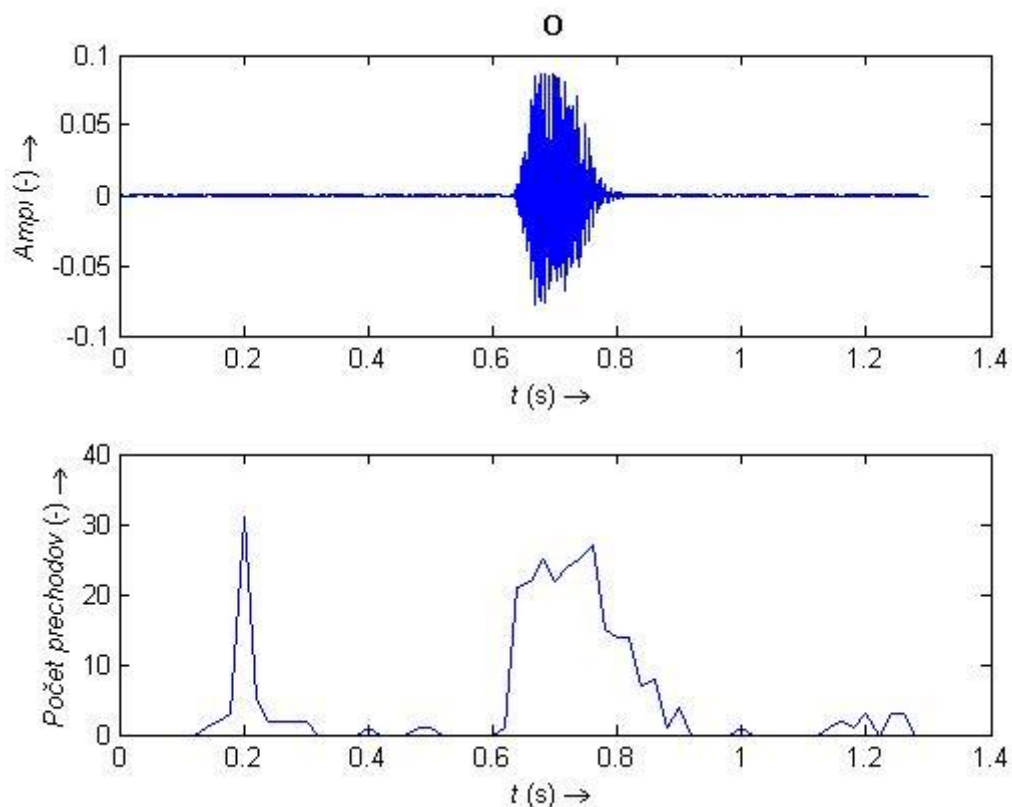
Obr. 4.4 Priebeh neznej spoluhlásky „c“ s vypočítaným počtom prechodov nulou



Obr. 4.5 Priebeh neznej spoluhlásky „s“ s vypočítaným počtom prechodov nulou



Obr. 4.6 Priebeh znej samohlásky „i“ s vypočítaným počtom prechodov nulou



Obr. 4.7 Priebek znelej samohlásky „o“ s vypočítaným počtom prechodov nulou

4.2.3 Krátkodobá autokorelačná funkcia

Využíva sa pri zisťovaní prítomnosti periódy v signále, teda na určovanie znelých a neznelých segmentov reči. Je definovaná vzťahom

$$R_n(m) = \sum_{k=-\infty}^{\infty} s(k)w(n-k)s(k+m)w(n-k-m), \quad (4.7)$$

kde $w(n)$ je okienko (pravouhlé alebo Hammingovo), $s(k)$ je signál v čase k . Ak je spracovávaný segment periodický, prejaví sa to zvýšením hodnôt autokorelačnej funkcie v bodoch $m=0, P, 2P, \dots$, kde P je perióda signálu. Práve túto vlastnosť možno využiť na určenie základnej periódy, ktorá bude na frekvencii P . Podmienkou pre správne fungovanie tejto metódy je, aby segment obsahoval aspoň dve periódy signálu. To je však dĺžkou okienka 20 ms zaručené pre frekvencie nad 100 Hz, avšak spracovávaný rečový signál môže mať základnú frekvenciu aj nižšiu, konkrétne od 70 Hz, preto sa používa dĺžka okienka 40 ms, ktorá pokrýva frekvencie od 50 Hz.

4.2.4 Metóda detekcie vrcholov

Táto metóda patrí medzi jednoduché metódy určovania periodických častí. Vychádza z predpokladu, že sa v znelom úseku reči opakuje celý priebeh periódy, teda aj vrcholy. Medzi nedostatky patrí závislosť na rečníkovi, časté vyhodnotenie nepresne

vypočítanej fundamentálnej frekvencie a možnosť vypočítať periódu len v špecifických bodoch – vrcholoch, nie v ľubovoľnom mieste signálu. Navyše je treba dbať na rozdielne amplitúdy signálu a preto meniť prah, ktorý určuje hranicu pod ktorou sú vrcholy zanedbané.

5 REALIZÁCIA APLIKÁCIE

Od pôvodného zámeru vytvoriť samostatnú aplikáciu sa nakoniec ustúpilo k použitiu už existujúceho nástroja, ktorý poskytuje žiadanú funkcionálnu a bolo teda možné naplno sa sústrediť na náplň práce a nie na implementáciu už existujúcich nástrojov. Preto bolo pre túto prácu použité prostredie programu Matlab, ktoré obsahuje nástroje nie len pre zložité výpočty, ale aj grafické rozhranie.

5.1 Dôvody použitia metód

Pre realizáciu algoritmu boli vybrané rôzne metódy schopné spracovávať rečový signál v časovej oblasti a ich kombinácie s rôznymi parametrami pre optimalizáciu výkonu a dosahovaných výsledkov. Na vstup bolo použité pravouhlé okienko s dĺžkou 20 ms, avšak toto okienko je často predĺžené na 40 ms, ako už bolo spomenuté. Dôvodom je krátkodobá autokorelačná funkcia s modifikáciami. Ďalším vybraným spôsobom spracovávania rečového signálu v časovej oblasti je metóda detekcie vrcholov. V niektorých častiach programu sa využíva ešte výpočet krátkodobej energie pre správne určenie základnej frekvencie z viacerých detekovaných hodnôt.

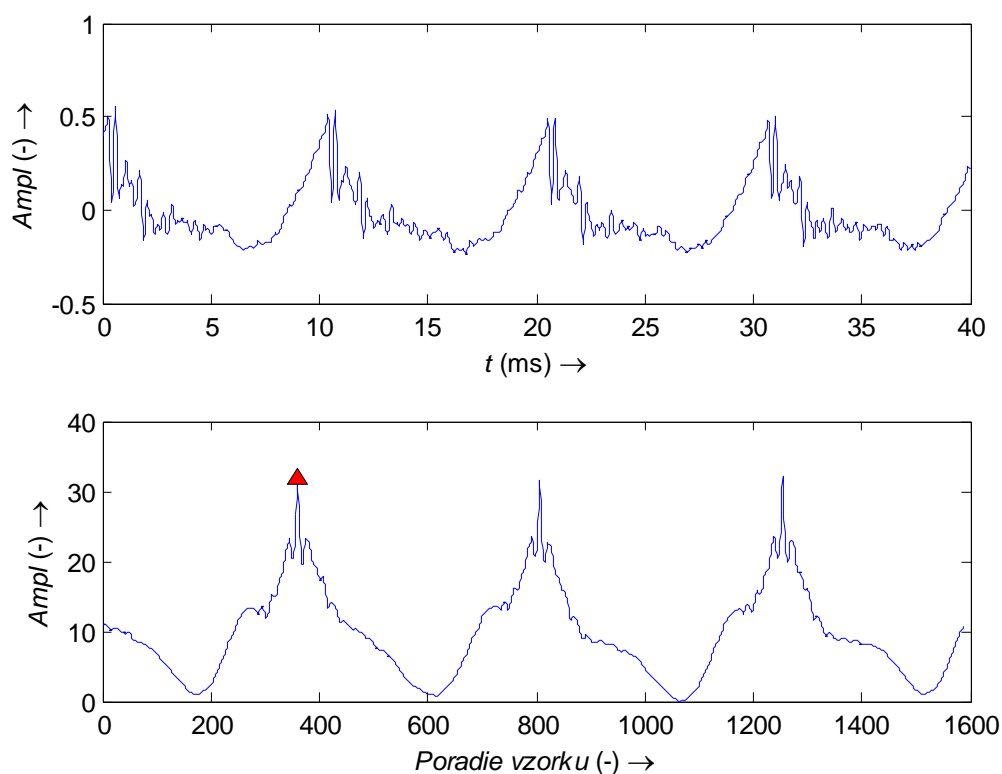
5.1.1 Krátkodobá autokorelačná funkcia

Metóda krátkodobej autokorelačnej funkcie je v mnohých zdrojoch považovaná za spoľahlivú metódu pri určovaní fundamentálnej frekvencie, čo je veľmi dôležitá podmienka pre ďalšie spracovanie signálu. Preto jej modifikácia bola zvolená za prvú metódu, aby mohli byť získané nové a nie už dosiahnuté výsledky v iných prácach, ktoré budú tiež naprieč širším spektrom signálov relatívne presné.

Porovnávaním malej časti segmentu – mikrosegmentu s celým segmentom s posunom po jednom vzorku vznikne nový upravený segment. Jednotlivé prvky tohto segmentu boli získané ako suma rozdielov všetkých prekrývajúcich sa prvkov zvoleného mikrosegmentu a pôvodného segmentu. Výsledný priebeh je skrátený o dvojnásobok zvolenej časti mikrosegmentu, pre aspoň čiastočné zníženie nárokov na výpočtovú techniku.

Zvolený mikrosegment je na začiatku segmentu a je teda rovnaký ako táto jeho časť a netreba ich porovnávať. Tento počet vzorkov sa však nesmie zanedbať v ďalšom spracovávaní. Rovnako sa nepočíta časť na konci segmentu, kde už presahuje mikrosegment do ďalšieho segmentu, pretože na konci signálu by mohlo dôjsť k porovnávaniu s neexistujúcou časťou signálu, program by sa pokúsil načítať hodnoty mimo povolenú oblasť v pamäti, čo spôsobí predčasné ukončenie programu.

Výstupom krátkodobej autokorelačnej funkcie je nový vyhladený priebeh signálu ako vidieť na obr 5.1, ktorého minimum prípadne v opačnom priebehu maximum znázorňuje koniec periódy. Výsledná perióda teda obsahuje počet prvkov, ktorý je súčtom pozície minima (maxima) a počtu prvkov v mikrosegmente.



Obr. 1.2 Priebeh signálu v predĺženom pravouhlom okienku (hore) a vyhladený priebeh signálu po spracovaní modifikovanou autokorelačnou funkciou (dolu) so zvýrazneným maximom

Nevýhodou autokorelačnej aj modifikovanej autokorelačnej funkcie sú nároky na výpočtovú techniku, ktoré rastú s dĺžkou zvoleného mikrosegmentu. Pri zvolení väčšej dĺžky mikrosegmentu musí program vykonať väčší počet výpočtov ako zobrazuje Tab.2 pre vzorkovaciu frekvenciu 44 100 Hz.

Tab. 2 Porovnanie požiadaviek na výpočtovú techniku autokorelačnou a modifikovanou autokorelačnou funkciou

Počet prvkov zvoleného mikrosegmentu	Počet vykonaných operácií s plnou dĺžkou	Počet vykonaných operácií so skrátenou dĺžkou	Zníženie výpočtov v % (údaje sú zaokrúhlene)
44	1 705 616	1 537 184	10
88	6 822 464	5 467 264	20
176	27 289 856	16 417 280	40
220	42 640 400	21 392 800	50

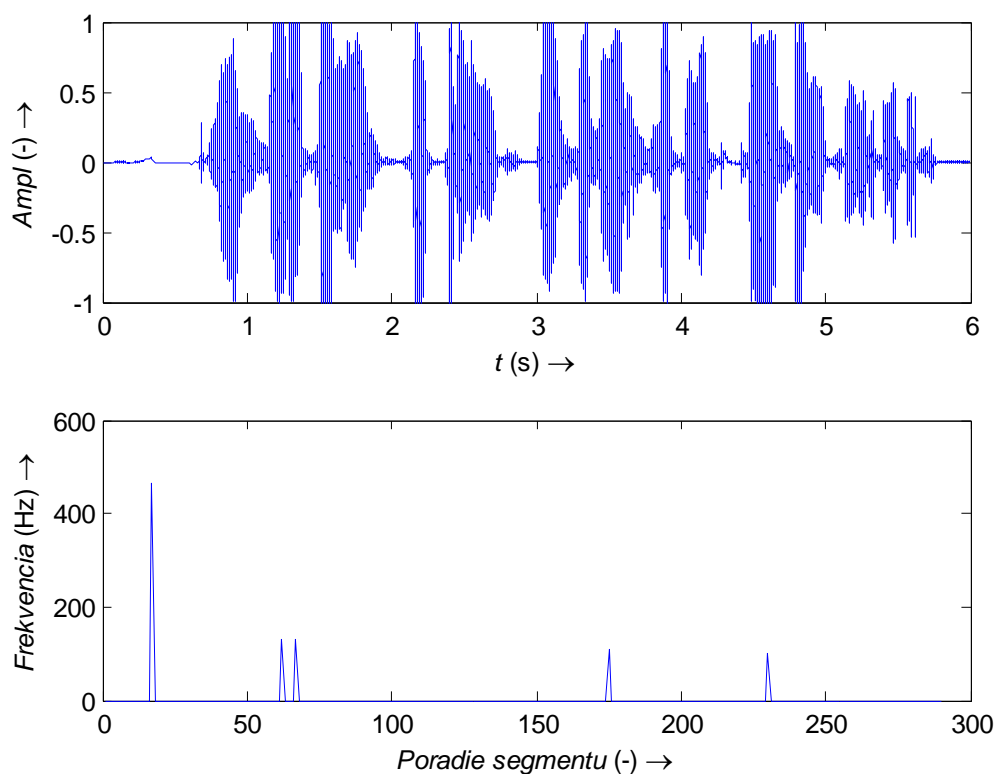
Porovnávaný úsek signálu si nemožno vybrať len na základe rýchlosti spracovania. Pri zvolení príliš dlhej časti signálu hrozí nebezpečenstvo, že vybraný mikrosegment už

obsahuje viac ako jednu periódu signálu. V tom prípade môže nastať nepresné určenie základnej periódy, čo má nezanedbateľný dopad na ďalšie spracovávanie rečového signálu, napríklad pri výpočte jitteru a shimmeru. Takto získané hodnoty možno považovať za nedôveryhodné a je potrebné použiť viac metód na kontrolu a porovnávanie získaných parametrov rečového signálu.

Samostatná autokorelačná metóda je navyše necitlivá na prechody – zmeny hlások, preto detekuje periódu aj v segmentoch, ktoré ju neobsahujú v celej časti – hlavne pre vyššie frekvencie, kde segment obsahuje viac periód a u autokorelačnej metódy sa vyhľadáva iba maximum, nie periodicky sa opakujúci priebeh.

5.1.2 Metóda detekovania vrcholov

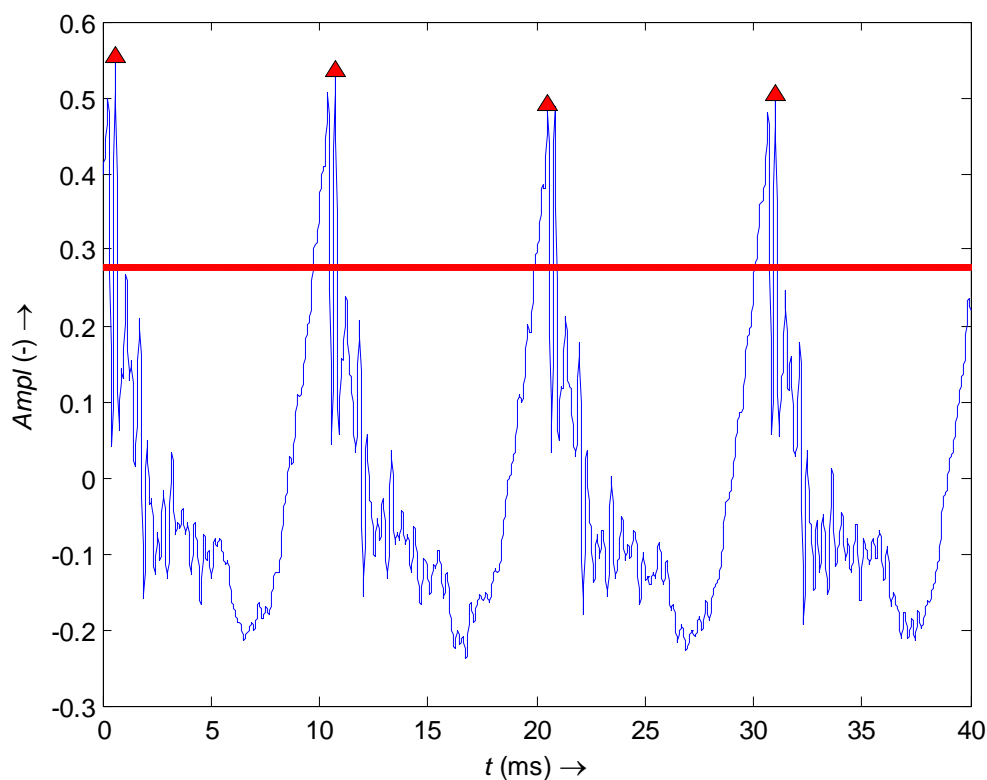
Detekovanie vrcholov je jednoduchý a na výpočtovú techniku nenáročný spôsob ako určiť základnú periódu. Použitím správnych parametrov, konkrétne prahu nad ktorým sa vrcholy nachádzajú a minimálnou vzdialenosťou vrcholov možno získať hodnotu základnej frekvencie.



Obr. 3.4 Celý signál (hore) so základnou frekvenciou okolo 100 Hz, vypočítaná základná frekvencia len pomocou metódy detekovania vrcholov (dolu)

Nevýhodou je, že perióda sa dá pomocou tejto metódy detekovať len vo vrcholoch, nie v ľubovoľnom mieste signálu, nakoľko nedbá na celý priebeh signálu na rozdiel od autokorelačnej funkcie. Navyše je nevhodná pre detekciu nízkych frekvencií ako možno vidieť na obr.5.2. Taktiež je často nájdená perióda nesprávna, hlavne ak sa v signále opakujú dva podobné vrcholy, ktoré si menia poradie vzhľadom na amplitúdu (obr.5.3).

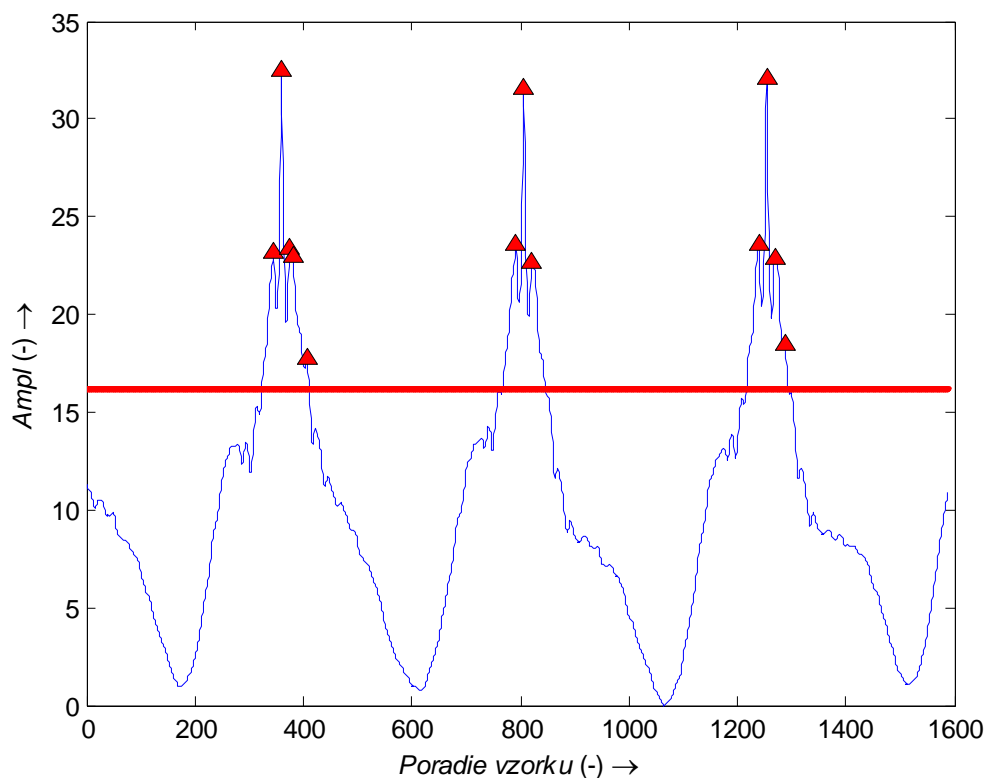
Navyše je závislá na rečníkovi, lebo zvolené parametre prahu a vzdialenosti vrcholov pre jeden subjekt nesplňujú podmienky pre správnu funkcionálnu spracovávajú audiozáznamov z databáze. Táto metóda by sa teda mohla javiť ako nepoužiteľná.



Obr. 5.6 Použitie detekcie vrcholov na reálnom signále, zachytená nepresnosť pri detekcii vrcholov v rečovom signále s nízkou základnou frekvenciou

5.1.3 Spoločná metóda

Na základe viacerých pokusov sa ukázalo, že pri zvolení premenného prahu a bez vzdialenosti vrcholov aplikovaných na výstupný priebeh z modifikovanej krátkodobej autokorelačnej funkcie je získaná perióda presnejšia ako len samotná autokorelačná metóda, hlavne rieši jej nedostatok s vyhľadáváním základnej periódy, teda detekuje periódu len v segmentoch, ktoré sa periodicky opakujú počas celého trvania segmentu. Využitím detekcie vrcholov sa výrazne nezvyšujú požiadavky na výpočtovú techniku.



Obr. 7.5 Výstup autokorelačnej metódy aplikovanej na dva segmenty so zvýrazneným prahom a detekovnými vrcholmi

Spoločná metóda je založená na výstupe modifikovanej autokorelačnej metódy s opačným priebehom, kde je najpodobnejší porovnávaný signál zobrazovaný ako lokálne maximum. Výstup je ešte posunutý tak, aby všetky hodnoty patrili medzi nezáporné reálne čísla. Dôvodom je správna funkcia vyhľadávania vrcholov, to znamená, že najpresnejšie hodnoty sú vo vrcholoch a v znelých segmentoch sa tieto hodnoty s podobnou amplitúdou približne periodicky opakujú. Výsledná perióda je teda definovaná len v prípade, že vrcholy sa približne periodicky opakujú a amplitúda týchto vrcholov sa výrazne nemení. Ako vidieť na obr. 5.4 je napriek použitiu autokorelačnej metódy, ktorá vyhladzuje priebeh, detekované veľké množstvo vrcholov. Prechádzajú sa všetky kombinácie vrcholov, aby medzi nájdenými hodnotami bola určite aj hodnota skutočnej fundamentálnej periódy. Prvá perióda je v tejto metóde vždy určená ako súčet vzdialenosti n -tého vrcholu a mikrosegmentu, druhá ako rozdiel vzdialeností m -tého a n -tého vrcholu, kde m a n je poradie vrcholov.

Napriek snahe získať len jednu a skutočnú hodnotu základnej periódy, je často získaných viac hodnôt. Preto použitím metódy krátkodobej energie na všetky vypočítané hodnoty periódy v jednom segmente sa rozhodne, ktorá perióda je najpravdepodobnejšie skutočná. Nakoľko obsahuje nie len fundamentálnu, ale aj jej násobky, ani táto kontrola nemusí vždy odhaliť správnu hodnotu. Preto bola použitá ešte jedna metóda, s inými vstupnými parametrami.

5.1.4 Kontrolná metóda

Ako naznačuje jej názov, bola zvolená na kontrolu už dosiahnutých výsledkov. Využíva výstup modifikovanej autokorelačnej funkcie, avšak nepoužíva opačné hodnoty, ani neposúva hodnoty medzi kladné. Rovnako ako spoločná metóda využíva detekciu vrcholov, kde však vrcholy nedefinujú najpodobnejšiu, ale najrozdielnejšiu časť priebehu zvoleného mikrosegmentu a segmentu. Vychádza z predpokladu, že nie len rovnaké, ale aj rozdielne časti sa periodicky opakujú.

5.1.5 Výsledná metóda

Táto metóda má dve realizácie. Buď je založená na výsledkoch spoločnej metódy a následne je priebeh skontrolovaný a prípadne aj zmenený na základe kontrolnej metódy alebo je naopak založená na kontrolnej metóde a na kontrolu je použitá spoločná metóda. Druhá opísaná možnosť bola aplikovaná po tom, čo sa ukázalo, že takto získané hodnoty môžu dosahovať lepšie výsledky, ako prvý spôsob realizácie.

5.2 Parametre aplikácie

Snahou tejto práce je vytvoriť aplikáciu, ktorá je nezávislá na rečníkovi, na kategórii rečového signálu (šepot, spev, čítanie...) a je schopná spracovávať dlhé rečové záznamy naprieč širokou databázou.

5.2.1 Objem vstupných dát

Aplikácia je prispôbená na spracovávanie rečového signálu so vzorkovaciou frekvenciou 44 100 Hz, 16 bitový mono alebo stereo záznam vo formáte waveform audio file - WAV. Na vstupe dokáže spracovať dlhý rečový signál (slová, vety, súvetia v rozsahu niekoľkých minút). Bola otestovaná na vzorkovacích frekvenciách do 384 kHz. Pri takto vysokej frekvencii už nie je schopná načítať súbor, pretože obsahuje priveľa hodnôt a dochádza k pretečeniu pamäti, čo spôsobí predčasné ukončenie. Preto je doporučené používať vzorkovaciu frekvenciu, na ktorú je prispôbená, prípadne pri vyššej znižovať dĺžku spracovaného rečového signálu.

5.2.2 Segment, mikrosegment, prah, vzdialenosť vrcholov

V tejto práci je využívaný segment s dĺžkou 20 ms, avšak pre základné frekvencie pod 100 Hz je predĺžený na 40 ms, aby obsahoval aspoň dve periódy signálu. Dĺžka mikrosegmentu je definovaná ako premenná o rozsahu 1/10 pôvodného segmentu, čo pri vzorkovacej frekvencii 44 100 Hz odpovedá 88 vzorkám. Takto dlhý mikrosegment pri využití modifikovanej autokorelačnej metódy redukuje využitie výpočtovej techniky o 20% bez vplyvu na rozsah detekovateľných základných frekvencií. Prah je vo všetkých metódach určený ako polovica maximálnej amplitúdy v danom segmente a vzdialenosť vrcholov je 1/10 pôvodného segmentu (88 vzoriek) pre samostatnú metódu detekcie vrcholov, pre iné metódy nebolo potrebné túto hodnotu definovať – môžu byť vzdialené jeden prvok.

5.2.3 Rozsah detekovateľných základných frekvencií

Detekovateľné základné frekvencie patria do rozsahu 50 až 500 Hz. To pokrýva všetky teoretické hodnoty fundamentálnej frekvencie pre mužov aj ženy, čiastočne deti. Tento rozsah sa dá rozširovať zmenou dĺžky mikrosegmentu a vzdialenosti vrcholov, avšak na úkor zníženej presnosti. Už v tomto rozsahu vidieť vplyv formantových frekvencií, ktoré sú občas nesprávne vyhodnotené ako základná frekvencia, vo väčšom rozsahu sa budú objavovať častejšie.

5.2.4 Jitter, shimmer

Jitter je zmena základnej frekvencie v dvoch za sebou idúcich periódach. Vypočíta sa teda ako rozdiel prvej a druhej periódy. Jeho hodnota je v Hz, prípadne sa prepočítava na percentá. Jeho hodnota sa vypočíta pomocou [9]

$$jitter(abs) = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{(i+1)}| \quad (5.1)$$

pre absolútnu hodnotu jitteru alebo [9]

$$jitter(rel) = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (5.2)$$

pre relatívnu hodnotu, kde N je počet hodnôt jitteru, T_i je i -tá hodnota základnej periódy.

Shimmer je kolísanie amplitúdy v dvoch za sebou idúcich periódach. Vypočíta sa ako rozdiel amplitúd. V tejto práci sa vypočítava v dvoch bodoch – v maximách a minimách periódy. Rovnako ako jitter sa prepočítava do percentuálneho vyjadrenia s referenčnou hodnotou maxima, respektíve minima prvej periódy podľa [9]

$$shimmer(dB) = \frac{1}{N-1} \sum_{i=1}^{N-1} |20 \log(A_{i+1} / A_i)| \quad (5.3)$$

pre absolútnu hodnotu v dB alebo [9]

$$shimmer(rel) = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |20 \log(A_i - A_{i+1})|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (5.4)$$

pre relatívnu hodnotu, kde N je počet hodnôt shimmeru a A_i je i -tá hodnota amplitúdy.

Hodnoty jitteru a shimmeru sú vypočítavané v absolútnej aj relatívnej hodnote, aby sa dali použiť v štatistických údajoch pre širšie spektrum rečníkov. Ich hodnoty by sa mali pohybovať hlavne v okolí 1% hodnôt vzhľadom na predchádzajúcu periódu. V tejto práci sa hodnoty jitteru a shimmeru vyhodnocujú pre všetky metódy, aby bolo možné porovnávať ich navzájom.

6 ŠTATISTICKÉ ÚDAJE

Táto časť je zameraná na porovnávanie spracovaných údajov jednotlivými metódami pre jeden rečový záznam alebo naprieč celou databázou záznamov. Navyše sú tu vypočítané základné štatistické údaje – stredná aritmetická hodnota a smerodajná odchýlka. Najprv sa vypočíta stredná aritmetická hodnota, teda aritmetický priemer podľa

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (6.1)$$

potom smerodajná odchýlka

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}, \quad (6.2)$$

kde N je počet hodnôt základnej periódy a x_i je i -tá hodnota základnej periódy.

Výstupom aplikácie je výpis neperiodických segmentov pre každú metódu, graf celého vstupného rečového signálu, grafy základnej periódy v závislosti na segmente pre každú metódu, absolútna a relatívna hodnota jitteru a shimmeru, histogramy jitteru v absolútnej hodnote a štatistické hodnoty – stredná aritmetická hodnota základnej frekvencie v spracovanom rečovom signále a smerodajná odchýlka tejto frekvencie.

6.1 Jeden rečový záznam spracovaný všetkými metódami

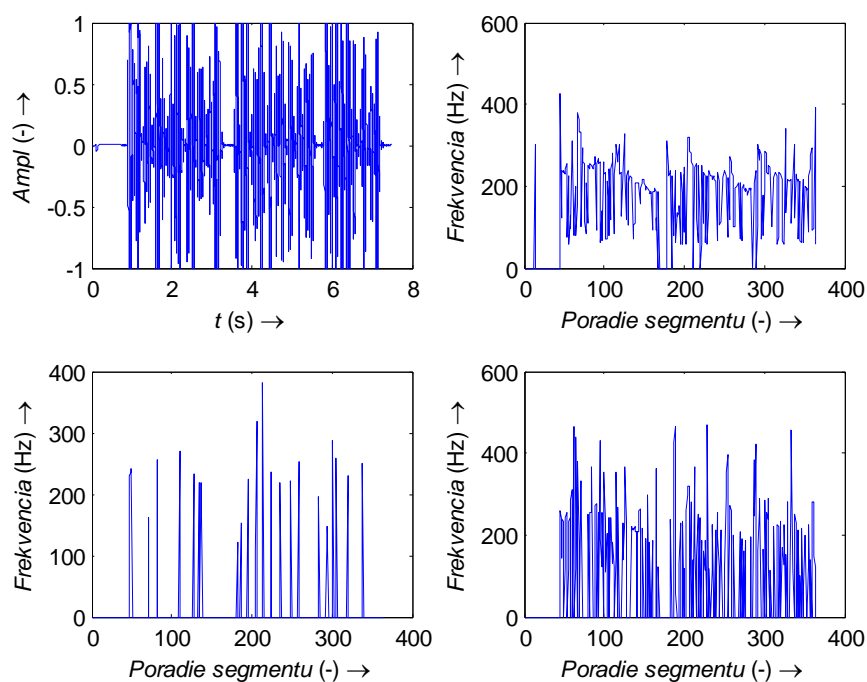
Ako vidieť v Tab. 3, samostatná autokorelačná metóda detekuje základnú frekvenciu v najväčšom počte segmentov. Je to však na úkor presnosti, pretože detekuje frekvenciu aj v segmentoch, v ktorých sa perióda neopakuje celý segment, preto sú aj hodnoty shimmeru najvyššie. Metóda detekcie vrcholov naopak detekuje veľmi malý počet segmentov s fundamentálnou frekvenciou. Napriek tomu má hodnota jitteru väčší rozptyl ako pri metódach s kontrolou.

Tab. 3 Vypočítané výstupné parametre aplikácie pre jeden rečový záznam

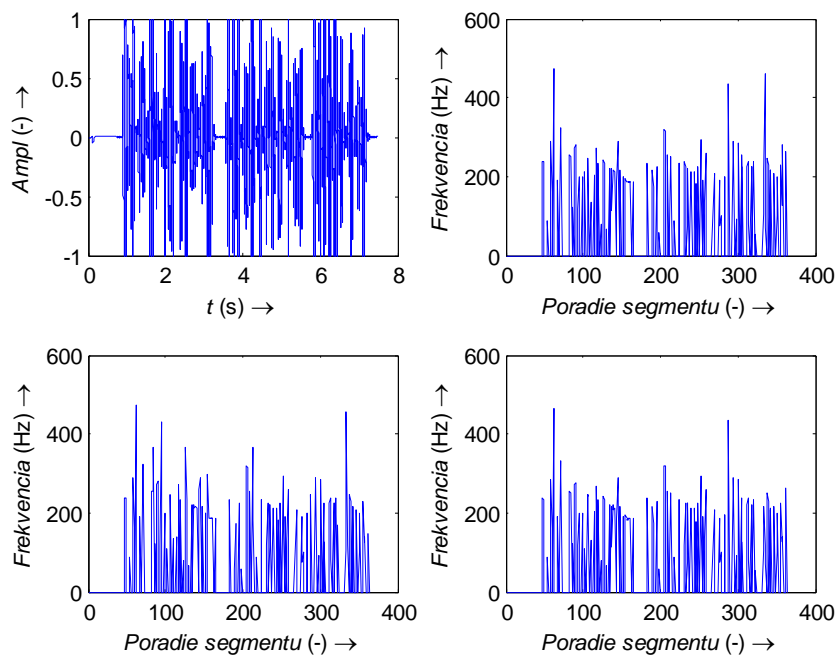
	Samostatná autokorelačná metóda	Metóda detekcie vrcholov	Spoločná metóda	Kontrolná metóda	Výsledná metóda na základe spoločnej	Výsledná metóda na základe kontrolnej
Počet segmentov s detekovanou základnou frekvenciou	305	28	163	118	163	118

Stredná aritmetická hodnota zákl. frekvencie (Hz)	195,32	220,01	229,98	211,10	210,03	203,58
Smerodajná odchýlka frekvencie	74,29	59,13	84,79	74,07	78,73	71,81
Počet hodnôt jitteru	614	58	551	293	286	286
Absolútna hodnota jitteru (Hz)	-0,75	-0,62	-0,70	-0,19	-0,46	-0,04
Relatívna hodnota jitteru (%)	0,34	0,32	0,39	0,10	0,24	0,02
Počet hodnôt shimmeru	614	58	551	293	286	286
Absolútna hodnota shimmeru (max) (dB)	5,23	1,01	5,14	2,92	2,63	2,68
Relatívna hodnota shimmeru (max) (%)	15,16	1,65	13,75	7,04	7,44	7,47
Absolútna hodnota shimmeru (min) (dB)	5,35	1,20	5,14	2,57	2,87	2,55
Relatívna hodnota shimmeru (min) (%)	15,85	3,14	14,33	6,73	7,44	7,29

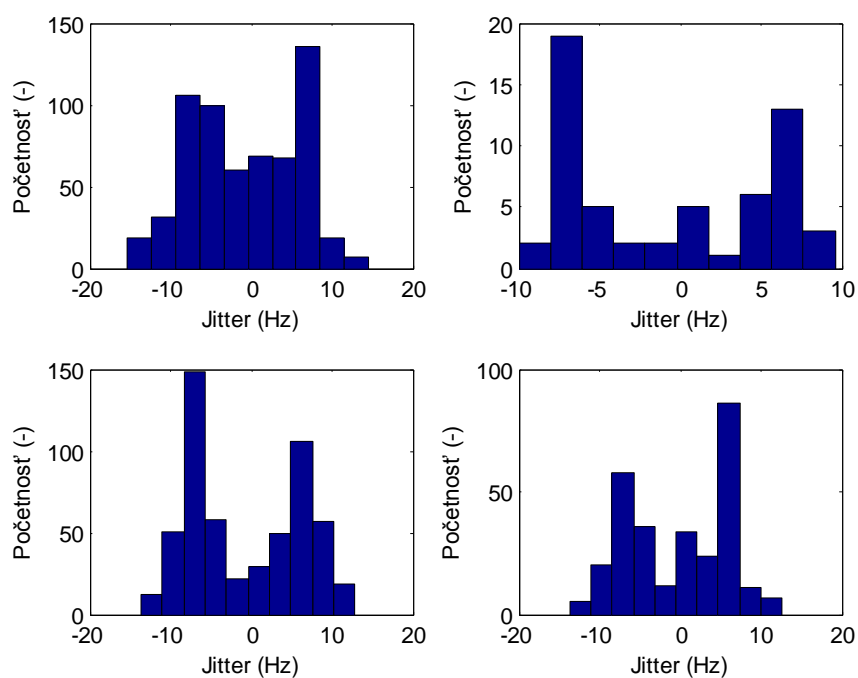
Nasledujúce grafy a histogramy sú výstupom aplikácie. Zobrazujú hodnoty základnej frekvencie vzhľadom na poradie segmentu. Grafy patria vždy k poslednému spracovanému rečovému signálu, histogramy sa vypočítavajú pre všetky spracované signály, až po ukončenie programu. Sú určené hlavne pre názornejšie porovnanie jednotlivých metód.



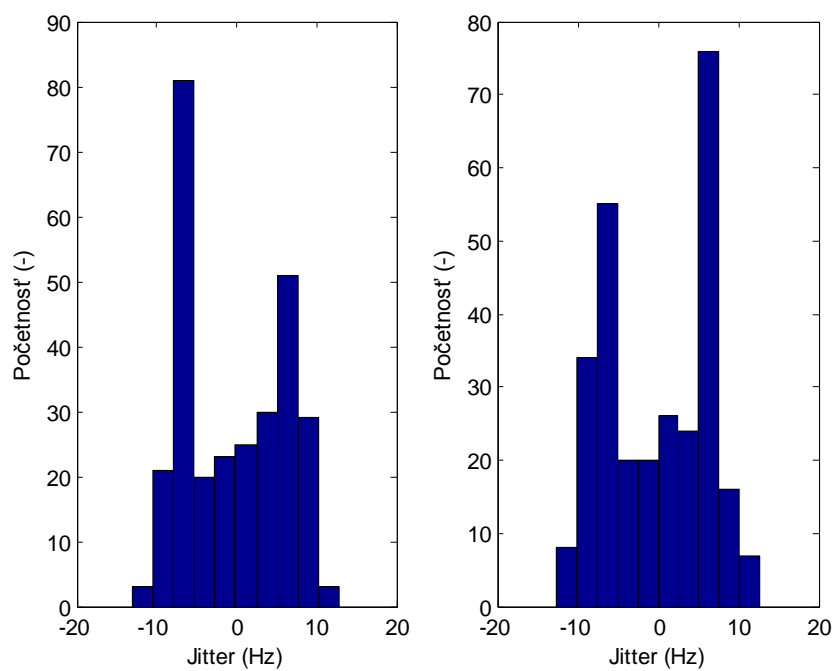
Obr. 1.2 Graf celého signálu (vľavo hore), hodnoty základnej periódy vzhľadom na poradie segmentu po spracovaní modifikovanou autokorelačnou metódou (vpravo hore), metódou detekcie vrcholov (vľavo dole) a spoločnou metódou (vpravo dole)



Obr. 3.4 Graf celého signálu (vľavo hore), hodnoty základnej periódy vzhľadom na poradie segmentu po spracovaní kontrolnou metódou (vpravo hore), výstupnou založenou na spoločnej (vľavo dole) a výstupnou založenou na kontrolnej (vpravo dole)



Obr. 5.6 Histogramy jitteru po spracovaní modifikovanou autokorelačnou metódou (vľavo hore), detekciou vrcholov (vpravo hore), spoločnou metódou (vľavo dole) a kontrolnou metódou (vpravo dole)



Obr. 7.6 Histogramy jitteru po spracovaní výslednou metódou založenou na spoločnej metóde (vľavo) a výslednou metódou založenou na kontrolnej metóde (vpravo)

6.2 Databáza záznamov spracovaných všetkými metódami

V tejto sekcii je venovaná pozornosť spracovávaniu viacerých záznamov naprieč databázou s rovnakým parametrom – rovnaký rečník, rovnaký text alebo rovnaká kategória reči.

6.2.1 Rovnaký rečník

Tab. 4 zobrazuje spracovanie viacerých záznamov rovnakého rečníka – ženy, v rôznych kategóriách reči. Tab. 5 zachytáva naopak mužský hlas s nižšou základnou frekvenciou. Sú zvolené výrazne odlišné frekvencie pre lepšiu možnosť porovnávania výstupných informácií.

Tab. 4 Vypočítané výstupné parametre aplikácie pre rovnakého rečníka (žena)

	Samostatná autokorelačná metóda	Metóda detekcie vrcholov	Spoločná metóda	Kontrolná metóda	Výsledná metóda na základe spoločnej	Výsledná metóda na základe kontrolnej
Počet segmentov s detekovanou základnou frekvenciou	1849	365	1280	1069	1280	1069
Stredná aritmetická hodnota zákl. frekvencie (Hz)	217,30	268,03	262,88	255,33	257,76	248,34
Smerodajná odchýlka frekvencie	94,82	69,20	90,54	89,34	89,00	89,46
Počet hodnôt jitteru	4530	1031	4719	3421	3603	3504
Absolútna hodnota jitteru (Hz)	-0,85	-0,33	-0,49	0,01	-0,37	0,00
Relatívna hodnota jitteru (%)	0,44	0,21	0,30	0,00	0,23	0,00
Počet hodnôt shimmeru	4530	1031	4719	3421	3603	3504
Absolútna hodnota shimmeru (max) (dB)	5,13	0,49	4,26	2,56	2,72	2,76
Relatívna hodnota shimmeru (max) (%)	9,49	1,67	7,48	5,17	5,58	5,63
Absolútna hodnota shimmeru (min) (dB)	4,71	0,74	3,66	2,29	2,67	2,67
Relatívna hodnota shimmeru (min) (%)	9,71	2,43	6,87	4,78	5,31	5,31

Tab. 5 Vypočítané výstupné parametre aplikácie pre rovnakého rečníka (muž)

	Samostatná autokorelačná metóda	Metóda detekcie vrcholov	Spoločná metóda	Kontrolná metóda	Výsledná metóda na základe spoločnej	Výsledná metóda na základe kontrolnej
Počet segmentov s detekovanou základnou frekvenciou	2066	72	1233	867	1233	867
Stredná aritmetická hodnota zákl. frekvencie (Hz)	132,04	148,35	176,19	153,36	163,50	152,83
Smerodajná odchýlka frekvencie	94,60	88,03	120,78	106,52	107,49	105,26
Počet hodnôt jitteru	5121	75	3695	1855	2149	2107
Absolútna hodnota jitteru (Hz)	-4,19	-3,35	-1,01	0,68	-0,97	0,48
Relatívna hodnota jitteru (%)	1,73	1,37	0,42	0,26	0,39	0,19
Počet hodnôt shimmeru	5121	75	3695	1855	2149	2107
Absolútna hodnota shimmeru (max) (dB)	10,18	0,70	11,09	6,71	7,38	7,56
Relatívna hodnota shimmeru (max) (%)	25,56	3,08	30,68	13,62	20,18	20,60
Absolútna hodnota shimmeru (min) (dB)	9,41	1,04	8,35	6,37	6,66	6,62
Relatívna hodnota shimmeru (min) (%)	29,35	6,39	25,39	17,19	20,69	20,23

Výsledky pre rovnakého rečníka sú veľmi výrazne ovplyvnené rečovým záznamom obsahujúcim šepot, ktorý je taktiež súčasťou databáze. Aj napriek tomu je vidieť, že hodnota jitteru sa pohybuje v očakávaných hodnotách.

6.2.2 Rovnaký text v zázname naprieč databázou

Pomocou záznamov s rovnakým textom, ale s rôznymi rečníkmi možno porovnať a posúdiť, ktorá metóda je najmenej závislá na rečníkovi. Podľa hodnoty jitteru možno vidieť, že ako najpresnejšia metóda sa javí výsledná metóda založená na kontrolnej metóde.

Tab. 6 Vybrané výstupné parametre po spracovaní záznamov s rovnakým textom

	Samostatná autokorelačná metóda	Metóda detekcie vrcholov	Spoločná metóda	Kontrolná metóda	Výsledná metóda na základe spoločnej	Výsledná metóda na základe kontrolnej
Počet segmentov s detekovanou základnou frekvenciou	1335	241	715	599	715	599
Počet hodnôt jitteru	3876	534	1981	1396	1446	1433
Absolútna hodnota jitteru (Hz)	-3,57	-0,67	-0,87	-0,29	-0,77	-0,07
Relatívna hodnota jitteru (%)	1,95	0,40	0,43	0,15	0,39	0,04

6.2.3 Rôzne rečové módy: čítanie, šepot, rozprávanie, spev

V tejto časti sú zobrazené vybrané štatistické údaje pre 4 rôzne rečové módy. Najdôležitejším parametrom pri ich porovnávaní je hodnota shimmeru. Vďaka nej možno posúdiť, že najmenšie kolísanie amplitúdy je všeobecne pri speve, rozprávanie a čítanie sa výrazne nelíšia a najväčšie je u šepotu, čo je očakávané správanie.

Tab. 7 Vybrané výstupné parametre po spracovaní 5 záznamov s rečovým módom čítanie

	Samostatná autokorelačná metóda	Metóda detekcie vrcholov	Spoločná metóda	Kontrolná metóda	Výsledná metóda na základe spoločnej	Výsledná metóda na základe kontrolnej
Počet hodnôt jitteru	4834	199	1699	1768	2003	1986
Relatívna hodnota jitteru (%)	1,06	0,55	0,18	0,09	0,18	0,04
Počet hodnôt shimmeru	4834	199	1699	1768	2003	1986
Relatívna hodnota shimmeru (max) (%)	26,89	2,04	26,76	14,29	19,17	19,33
Relatívna hodnota shimmeru (min) (%)	26,88	8,04	22,98	16,29	19,22	19,19

Tab. 8 Vybrané výstupné parametre po spracovaní 5 záznamov s rečovým módom šepot

	Samostatná autokorelačná metóda	Metóda detekcie vrcholov	Spoločná metóda	Kontrolná metóda	Výsledná metóda na základe spoločnej	Výsledná metóda na základe kontrolnej
Počet hodnôt jitteru	4163	55	2732	1303	1590	1574
Relatívna hodnota jitteru (%)	3,10	5,37	0,56	0,04	0,47	0,01
Počet hodnôt shimmeru	4163	55	2732	1303	1590	1574
Relatívna hodnota shimmeru (max) (%)	47,52	4,58	33,38	31,90	31,43	32,18
Relatívna hodnota shimmeru (min) (%)	45,07	12,11	36,12	29,66	31,12	30,88

Tab. 9 Vybrané výstupné parametre po spracovaní 5 záznamov s rečovým módom rozprávanie

	Samostatná autokorelačná metóda	Metóda detekcie vrcholov	Spoločná metóda	Kontrolná metóda	Výsledná metóda na základe spoločnej	Výsledná metóda na základe kontrolnej
Počet hodnôt jitteru	6243	315	4888	2176	2705	2665
Relatívna hodnota jitteru (%)	1,41	1,03	0,28	0,14	0,24	0,09
Počet hodnôt shimmeru	6243	315	4888	2176	2705	2665
Relatívna hodnota shimmeru (max) (%)	30,56	4,68	31,88	17,84	21,04	20,90
Relatívna hodnota shimmeru (min) (%)	29,74	29,33	25,46	17,42	20,23	20,40

Tab. 10 Vybrané výstupné parametre po spracovaní 5 záznamov s rečovým módom spev

	Samostatná autokorelačná metóda	Metóda detekcie vrcholov	Spoločná metóda	Kontrolná metóda	Výsledná metóda na základe spoločnej	Výsledná metóda na základe kontrolnej
Počet hodnôt jitteru	3799	652	4022	2423	2685	2702
Relatívna hodnota jitteru (%)	0,92	0,38	0,34	0,04	0,15	0,06
Počet hodnôt shimmeru	3799	652	4022	2423	2685	2702
Relatívna hodnota shimmeru (max) (%)	12,60	1,68	11,53	7,22	8,70	9,06
Relatívna hodnota shimmeru (min) (%)	14,04	3,75	11,97	7,98	9,81	9,71

7 ZÁVER

V práci sú uvedené základné publikácie, ktoré boli zdrojom informácií nutných k oboznámeniu s touto problematikou. Taktiež sú podrobne rozoberané vybrané metódy spracovávania rečových signálov v časovej oblasti a ich použitie na reálnom časovom priebehu rečového signálu.

V ďalšej časti práca obsahuje ukážky niekoľkých vybraných metód aplikovaných na reálny rečový signál. Výsledná aplikácia však využíva len poznatky pre autokorelačnú metódu, metódu detekcie vrcholov a čiastočne využíva metódu krátkodobej intenzity. Táto metóda, ale nie je použitá na spracovávanie celého signálu, ale len na výstup z autokorelačnej metódy, respektíve na kontrolu správnej hodnoty základnej periódy.

Samotná aplikácia je realizovaná v prostredí Matlab, pretože toto prostredie podporuje formát wav, v ktorom je vytvorená celá databáza záznamov, rovnako je schopné zobrazovať výsledné hodnoty v grafoch, histogramoch, či príkazovom okne, čo aplikácia využíva. Navyše je schopné interaktívne komunikovať s užívateľom formou dialógových okien. Ďalším dôvodom použitia už vytvoreného prostredia je úspora času, ktorý mohol byť využitý k lepšej funkcionalite, napríklad na tvorbu kontrolnej metódy.

Štatisticky spracované výsledky majú veľkú smerodajnú odchylku. Táto hodnota nie je veľmi prekvapivá, nakoľko bolo snahou získať základnú periódu pre čo najväčší počet segmentov, pri čom sa nezanedbával signál s nízkou amplitúdou, kde sa prejavuje vplyv šumu.

Neočakávaným záverom je, že najlepšie hodnoty jitteru dosahuje po spracovaní viacerých rečových záznamov výsledná metóda, ktorá je založená na kontrolnej metóde. Táto metóda pôvodne nebola vôbec plánovaná ako výstupná, pretože využíva najväčší rozdiel priebehov mikrosegmentu a segmentu a vzdialenosť takto vzniknutých vrcholov považuje za možnú periódu. Dôvodom najväčšej presnosti je pravdepodobne fakt, že kontrolná metóda vypočítava veľké množstvo možných hodnôt základnej periódy a následne sú tieto hodnoty porovnávané na základe rozdielov intenzity. Frekvencia s najmenším rozdielom intenzity je považovaná za základnú periódu v danom segmente.

Ďalším prekvapivým výsledkom sú relatívne hodnoty shimmeru, ktoré sa podľa teoretických predpokladov majú pohybovať v rozsahu jedného percenta, avšak výstupné hodnoty shimmeru sa výrazne odlišujú. Najmarkantnejší rozdiel možno vidieť u spracovávania šepotu, kde relatívna hodnota shimmeru v niektorých metódach presahuje 50 %. Táto odchylka môže byť spôsobená príliš veľkým kolísaním amplitúd v maximách a minimách amplitúdy.

Naopak hodnoty jitteru takmer vždy splňujú teoretické očakávania, pre výsledné metódy, či už založené na spoločnej alebo kontrolnej metóde, sú nezávislé na rečníkovi, na móde reči ani na počte spracovaných záznamov.

Výsledky tejto bakalárskej práce splňujú teoretické predpoklady len pre niektoré parametre. Pre splnenie všetkých predpokladov by však bolo treba použiť ako základnú metódu inú, avšak pravdepodobne žiadna z dnes využívaných metód nedokáže dokonale

fungovať za každých možných podmienok.

LITERATÚRA

- [1] PSUTKA, J., MÜLLER, Z., MATOUŠEK, J., RADOVÁ, V. Mluvíme s počítačem česky. Praha: Academia, 2006.
- [2] SIGMUND, M. Rozpoznávání řečových signálů. Skriptum. Brno: FEKT, VUT v Brně, 2007
- [3] Internetový zdroj dostupný z :
<http://blog.doucovanie-vzdelavanie.sk/65/kde-sa-vzala-ludska-rec/>
- [4] Internetový zdroj dostupný z:
<http://www.lsg.sk/~kucera/Projekt/telo/dychanie.htm>
- [5] Internetový zdroj dostupný z:
http://www.fpv.umb.sk/kat/kf/Integrovana_prirodoveda/SG/Zvuk/Piskat/index1.htm
- [6] Internetový zdroj dostupný z:
<http://is.muni.cz/elportal/estud/ff/js07/fonetika/materialy/index.html>
- [7] Internetový zdroj dostupný z:
http://xeds.eu/other/P_Boersma_Accurate_short-term_analysis_of_the_fundametnal_freq.pdf
- [8] KÁŇA, L. Elektroakustika. Skripta FEKT VUT, Brno 2002
- [9] Internetový zdroj dostupný z:
http://nlp.lsi.upc.edu/papers/far_jit_07.pdf

ZOZNAM SYMBOLOV, VELIČÍN A SKRATIEK

F_0	Základná hlasová frekvencia, označovaná aj fundamentálna
F_{vz}	Frekvencia vzorkovania
FFT	Fast Fourier Transform, rýchla Fourierova transformácia

PRÍLOHA A

Obsah CD

- Zdrojové kódy aplikácie
- Databáza rečových záznamov vo formáte wav